ATLASES ARE ALREADY INSIDE: RECOVERING POPULATION TEMPLATES FROM PRE-TRAINED DIFFUSION MODELS

Anonymous authors
Paper under double-blind review

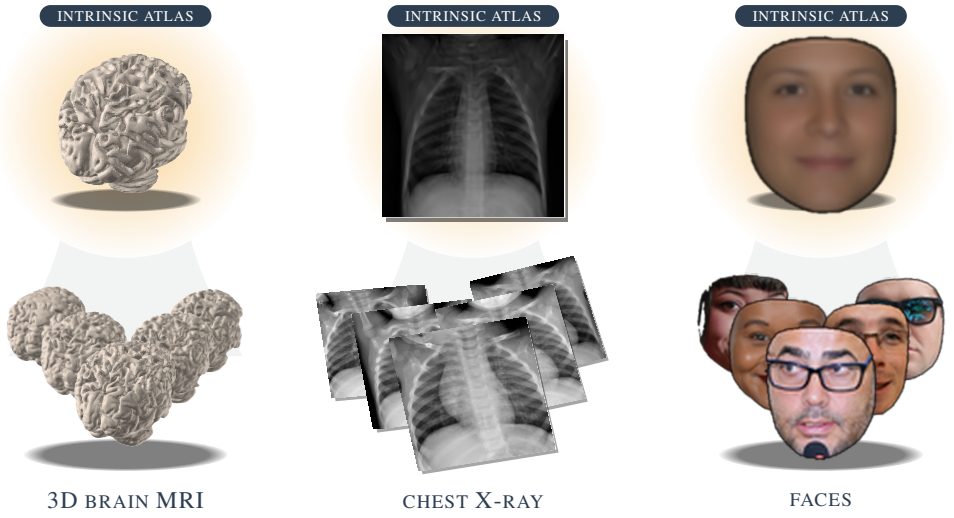


Figure 1: **Recovery of a population's central anatomy from diffusion models**, across different modalities, including 3D brain MRI, 2D chest X-rays, and 2D faces. Generators are trained to synthesize individuals (bottom). Our method extracts its intrinsic atlas from each generator (top).

ABSTRACT

We present a new inference-time sampler for diffusion models that gives a pre-trained model a capability it was never trained for: constructing the atlas of the population it synthesizes. The sampler converges from every random seed to the population's central anatomy, which we call the *intrinsic atlas*. The advantage is threefold. (1) It requires no retraining. A diffusion model that has already learned a coherent population, including the released ones, yields its atlas in a single inference pass without involving deformable registration. (2) It applies to multiple domains, such as brain MRI, chest X-ray, faces, and 3D shapes. (3) It extends to subpopulations. One age-conditioned model gives an atlas at any age in its training range, and the resulting family reproduces the CSF expansion of healthy aging. Evaluated as a registration target, the intrinsic atlas is best or second-best on every dataset against classical and learned templates, and the most central template on held-out brain MRI cohorts. Atlas construction can be reframed as a byproduct of generative modeling: a diffusion model is a learned representation of population structure, and the atlas is what it already contains.

1 INTRODUCTION

Every coherent population has a center, where every member inside it is a smooth deformation of one shared structure: the brain that all brains are variations of, the face behind all faces, the letter beneath every typeface. Grenander & Miller (1998) formalized this center as a template, in which

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

a population is one template together with the transformations that carry it to each member. This template is also known as an atlas in medical imaging. In practice, fields that need an atlas build it laboriously by spatially aligning thousands of individuals into a common frame under a chosen deformation model (Klein et al., 2009; Xu & Niethammer, 2019; Ding & Niethammer, 2022; Ranem et al., 2024). Learning-based methods amortize alignment into a network (Balakrishnan et al., 2019; Dey et al., 2021; Dalca et al., 2019; Yang et al., 2022; Abulnaga et al., 2025), but a deformation model remains what defines the atlas. In contrast, we show that the atlas can be obtained by a new sampling strategy on diffusion models (Ho et al., 2020; Song et al., 2021) trained for synthesis alone. This sampler converges from every random start to one sharp image (cross-seed SSIM ≥ 0.99), on brains, faces, and 3D shapes alike. We refer to this behavior as *canonical convergence*.

Our hypothesis behind it comes from a correspondence between the diffusion model and pattern theory. Pattern theory, as used in computational anatomy, decomposes a population into a template and the transformations that deform it to each member (Grenander, 1970; Grenander & Miller, 1998). The reverse process of a diffusion model has the same two terms. A learned denoiser moves the state toward the posterior mean of the population given the current noisy state, and an injected noise decides where within the population the trajectory goes. Prior work (Meng et al., 2021; Huberman-Spiegelglas et al., 2024) exploits the second term by editing the injected noise to edit samples, showing that the noise carries individual variation within the population. Our hypothesis concerns the complementary term. We posit that the denoiser carries what the population shares, namely, the template. If so, then for a coherent population, whose members share a single structure and differ only by continuous variation, the denoiser alone should trace out the population's central instance, which we call the *intrinsic atlas* in this work. The hypothesis makes three commitments, and each is testable. First, canonical convergence must be a property of the sampler, not of one network. Second, it must serve as a valid registration target. Third, it must report the structure of the population it was trained on, returning one image where the population has one template, one image per template where it has several, and no stable image where it has none. The rest of this paper tests all three.

The first commitment holds for every population that shares a central anatomy, medical and non-medical, in 3D and 2D, with a mechanism investigation in Section 3. For the second, we evaluate the intrinsic atlas as a registration target under a standard registration protocol, first on T1-weighted brain MRI, where atlas construction and its evaluation are most mature, then on faces, chest X-ray, and 3D shapes. With no registration, fine-tuning, or atlas objective, it is best or second-best on every dataset against classical and learned templates, and the most central and most regular template on held-out brain cohorts (Section 4). The template also moves with the generator's conditioning, where an age-conditioned model returns a family of age-specific brain atlases on demand that reproduces the CSF expansion of healthy aging (Section 5). For the third, populations with several templates or none resolve into multiple templates or into a blurry image (Section 6). Our contributions are as follows:

- **Canonical convergence by design.** A deterministic variant of sampling turns any diffusion model of a coherent population into an atlas constructor. We explain the convergence as a contraction of the noise-free operator, and verify it with controlled experiments.
- **Cross-domain registration benchmark.** Under a fixed registration protocol per domain, the intrinsic atlas is best or second-best across multiple domains, such as brain MRI, faces, chest X-ray, and 3D shapes. Our method is the only method that spans them all.
- **Conditioning and population structure.** The atlas inherits the generator's conditioning, giving an age-specific atlas family on demand. Where a population has several centers or none, the same sampler makes that visible.

2 Related Work

2.1 Atlas and template construction

Atlases in the medical field. Classical atlas construction alternates registration and averaging under an explicit deformation model, *e.g. LDDMM* (Beg et al., 2005) and unbiased-template diffeomorphic construction (Fonov et al., 2011; Klein et al., 2009; Tustison et al., 2021b). Learning-based methods amortize this optimization into a network. *VoxelMorph* (Balakrishnan et al., 2019) and follow-ups learn pairwise or groupwise registration. *Atlas-GAN* (Dey et al., 2021) and conditional

deformable-template networks (Dalca et al., 2019; Rakic et al., 2025) learn a network whose output is the atlas, trained under a template-construction or shape-reconstruction loss. *ImplicitAtlas* (Yang et al., 2022) represents the template implicitly. *MultiMorph* (Abulnaga et al., 2025) trains a feedforward model that produces a population-specific atlas in one forward pass, reporting a $100\times$ speedup over iterative construction. These methods were developed and validated on brain MRI, on manually labeled cohorts with an agreed protocol (Klein et al., 2009), which is the most mature benchmark setting for atlas construction. All of these construct the atlas with atlas-specific objectives. In contrast, we obtain the atlas from a generator trained for synthesis alone, without any atlas objectives.

Learned templates in vision. Outside medicine, joint-alignment methods optimize an aligner together with a template of an image collection: congealing averages the aligned stack (Learned-Miller, 2006), GANgealing aligns images to a GAN sample at a learned latent (Peebles et al., 2022), and neural congealing maps images into a joint atlas of semantic features (Ofri-Amar et al., 2023). For 3D shapes, deep implicit templates learn one implicit surface per category together with a per-shape deformation onto it, DIT through a learned warp field (Zheng et al., 2021) and DIF-Net through a deformation-plus-correction field (Deng et al., 2021). In every case, the template is a trained object, and its quality is measured under the aligner it was trained with. We train nothing for the template, and we evaluate it under a registration protocol it has never seen, alongside the templates these methods produce.

2.2 Deterministic sampling and the role of noise

Deterministic samplers such as DDIM and the probability-flow ODE (Song et al., 2020; 2021) are designed to preserve the forward marginals, so each initialization still yields a distinct sample. Previous work (Meng et al., 2021; Huberman-Spiegelglas et al., 2024) edits the injected noise during the sampling steps to edit a sample, exploiting the fact that the noise carries individual variation. Our sampler is the opposite operation that discards the noise rather than steering it, and, unlike DDIM, it does not preserve the marginals. The identification of the denoiser with the posterior mean and the score is classical (Efron, 2011). To our knowledge, no prior work reports this exact convergence, or its use as an atlas.

3 The Posterior-Mean Iteration and Canonical Convergence

We use a trained denoising diffusion probabilistic model (Ho et al., 2020), with forward process $x_t = \sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1-\bar{\alpha}_t}\,\epsilon$, $\epsilon \sim \mathcal{N}(0, I)$, and a network $\epsilon_\theta(x_t, t, y)$ trained to predict ϵ under an optional condition y (*e.g.* anatomy class or age). We do not modify training in any way; the only change is how the reverse process is run at inference.

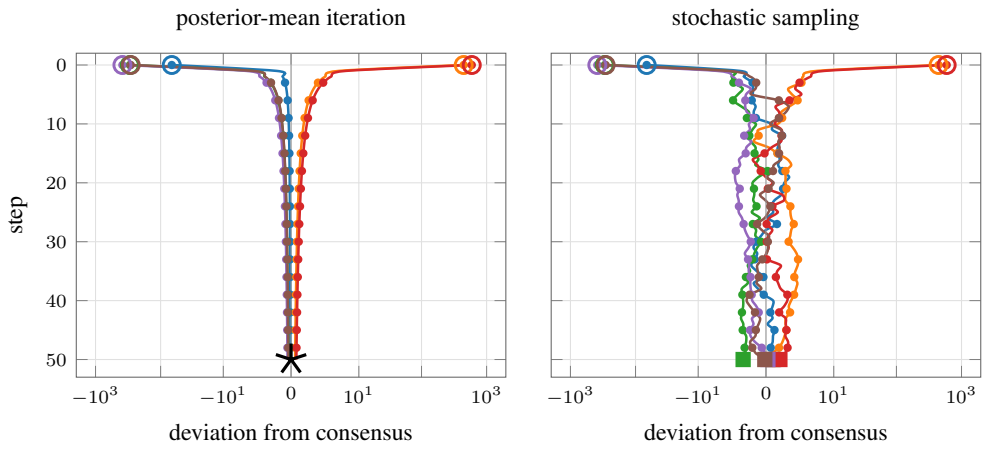


Figure 2: **Canonical convergence**. Six random initializations (circles) under the posterior-mean iteration contract onto one image (star) within the first steps; the same network sampled stochastically keeps them distinct (squares). Brain-T1 generator, first 50 of 900 steps.

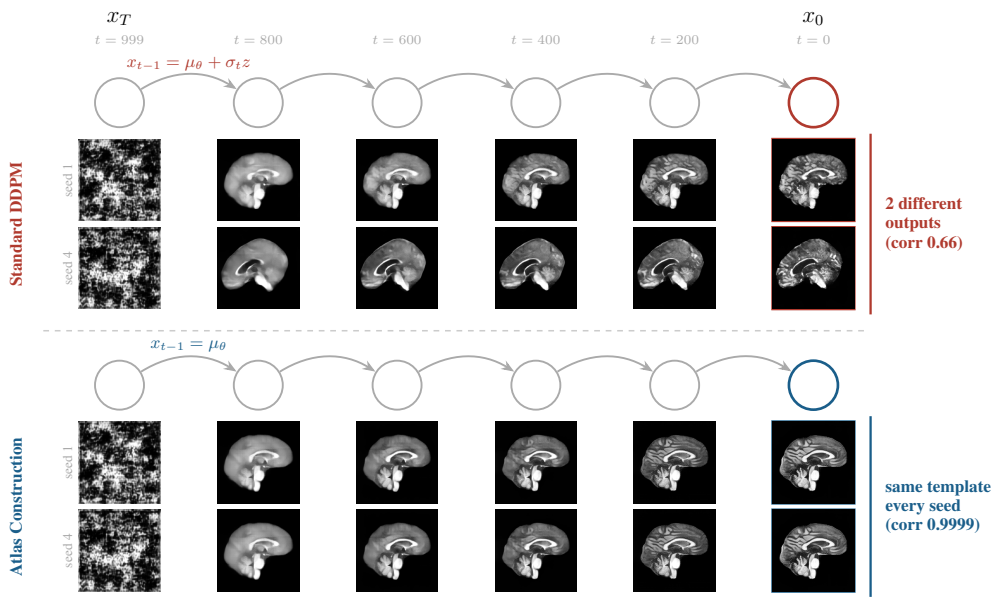


Figure 3: **Canonical convergence.** Reverse process of one pretrained brain-MRI diffusion model at six timesteps, decoding $\hat{x}_0 = \mathbb{E}[x_0 \mid x_t]$; each row uses an independent seed. **Top**: stochastic sampling, each seed a different brain (cross-seed correlation 0.66). **Bottom**: the posterior-mean iteration on the same network; every seed converges to the same image (0.9999).

The posterior-mean iteration. Rather than sampling $p_\theta(x_{t-1} \mid x_t, y)$, we replace each stochastic transition by its conditional mean,

$$x_{t-1} \leftarrow \mu_\theta(x_t, t, y) = \mathbb{E}_{p_\theta(x_{t-1}|x_t,y)}[x_{t-1}], \tag{1}$$

where ordinary sampling would add $\sigma_t z$, $z \sim \mathcal{N}(0, I)$, to this mean at every step. This differs from DDIM, which re-inserts the predicted noise at the schedule's amplitude, $x_{t-1} = \sqrt{\bar{\alpha}_{t-1}}\,\hat{x}_0 + \sqrt{1-\bar{\alpha}_{t-1}}\,\hat{\epsilon}$, and so preserves each initialization; the posterior-mean iteration re-inserts nothing, so the state is driven only by the denoiser's pull toward the population's center.

The mean update is determined by the denoised estimate $\hat{x}_0(x_t, t, y)$, itself linked to the score of the noised marginal via Tweedie's formula (Efron, 2011),

$$\hat{x}_0(x_t) = \tfrac{1}{\sqrt{\bar{\alpha}_t}}\big(x_t + (1-\bar{\alpha}_t)\nabla_{x_t} \log f_{X_t}(x_t)\big). \tag{2}$$

$\hat{x}_0(x_t) = \mathbb{E}[x_0 \mid x_t]$ is the expected clean image given the noisy state x_t. Without any injected noise, each update moves the state toward $\hat{x}_0$, the center of the population, from the current state. Figure 4 makes this concrete on a 2D Gaussian population, whose center and noised density f_{X_t} are known. With a known score function $\nabla_{x_t} \log f_{X_t}(x_t)$ in place of a trained network, all 300 seeds reach the center (offset 2×10^{-4}), while stochastic sampling from the same seeds spreads across the population. Convergence to the center is a property of the operator, not of a trained network.

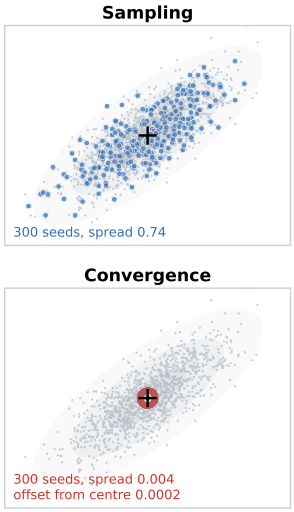
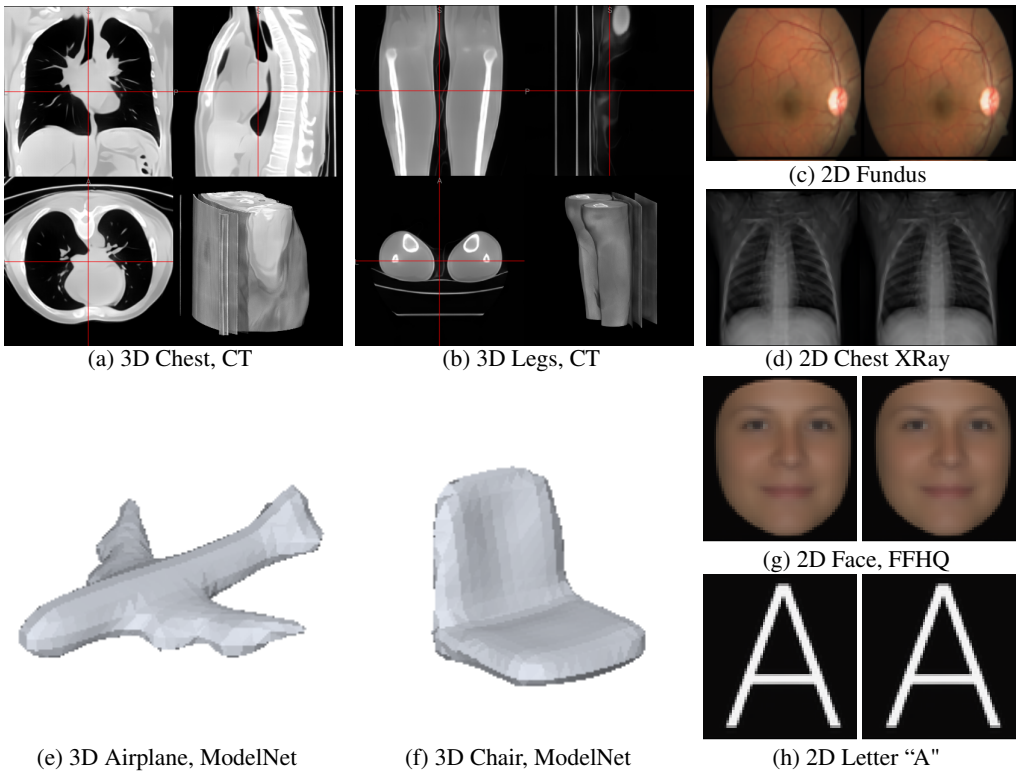


Figure 4: **Sampling versus convergence** on a population with a known center.

Canonical convergence and the intrinsic atlas. Let $\Phi_\theta(x_T; y)$ denote the endpoint obtained by running the posterior-mean iteration from the initialization x_T under condition y; it is a deterministic function of x_T. We say the model exhibits *canonical convergence* at y if $\Phi_\theta(x_T; y)$ is the same for almost every $x_T \sim \mathcal{N}(0, I)$; that common endpoint, written $x^\star(y)$, is the model's *intrinsic atlas* for condition y. This behavior can be verified easily: if a population center exists, every random x_T should return the same image; if not, the trajectories should scatter. As shown in Figure 2, comparing against the stochastic sampling, the posterior-mean iteration presents clear convergence.



(a) 3D Chest, CT (b) 3D Legs, CT (c) 2D Fundus (d) 2D Chest XRay (e) 3D Airplane, ModelNet (f) 3D Chair, ModelNet (g) 2D Face, FFHQ (h) 2D Letter “A"

Figure 5: **Generality.** Intrinsic atlases on medical (top) and non-medical (bottom) populations, in 3D and 2D: (a) chest CT, (b) leg CT, (c) retinal fundus, (d) chest X-ray, (e) ModelNet airplane, (f) ModelNet chair, (g) FFHQ faces, (h) the letter “A” across typefaces.

Atlas Visual Quality Across Modalities and Domains. As shown in Figure 5, we present intrinsic atlases from multiple domains, including chest and leg CT, retinal fundus, CelebA faces, the letter “A” across 3,796 typefaces, and ModelNet airplanes and chairs, each with cross-seed SSIM ≥ 0.99. The atlas keeps what the population agrees on and drops what it does not: the “A” is a clean sans-serif glyph from a mix of sans, serif, and script faces. Interestingly, the chair atlas has a seat and backrest but no legs, since the population does not share a leg design.

4 The Utility of the Intrinsic Atlas

Atlas construction has been developed and benchmarked almost entirely on brain MRI: the classical methods (Joshi et al., 2004; Avants et al., 2008; Fonov et al., 2011), the reference evaluation of registration algorithms (Klein et al., 2009), and every learned-template method we compare against (Dalca et al., 2019; Dey et al., 2021; Ding & Niethammer, 2022; Abulnaga et al., 2025) were designed and validated there. We therefore adopt this well-benchmarked domain as our primary evaluation, on IBSR18 (BrainFacts/SfN, 2007), Mindboggle101 (Klein & Tourville, 2012), and ABIDE-I (Di Martino et al., 2014), following its established protocols (Klein et al., 2009; Dalca et al., 2019; Abulnaga et al., 2025; Hering et al., 2022). Meanwhile, we evaluate on domains for template correspondence, namely CelebA (Liu et al., 2015) faces, Montgomery (Jaeger et al., 2014) chest X-rays, and 3D KeypointNet (You et al., 2020) chairs and cars. Since those domains have no anatomical segmentations, we report the percentage of correct keypoints (PCK) and the mean/median label transfer IoU, following (Peebles et al., 2022; Deng et al., 2021). Details are provided in the supplementary material.

Registration Utility of the Intrinsic Atlas. The intrinsic atlas is best or second in every dataset on each domain. On brain MRI, MultiMorph has a higher Dice by at most half a point on Mindboggle101 and ABIDE ($+0.004$ each), and by three points on IBSR18, while ours is the most central and most regular template on every cohort. The same best or second trend persists in other domains.

Table 1: **Atlas construction evaluation across domains.** Registration is ANTs in every domain, with one preset per domain (SyNQuick for brain MRI, SyNCC for faces and X-ray, SyN for shapes). Within a domain, the same registration protocol is used for every template, so only the template varies. Every cell is the mean of at least 3 repeated runs.

(a) **On T1 Brain MRI. Metrics**: We report Dice, the accuracy criterion of the reference benchmark (Klein et al., 2009), plus three deformation metrics such as centrality (Dalca et al., 2019; Abulnaga et al., 2025), mean per-subject displacement $|u|$ (Dalca et al., 2019), and SDlogJ (Hering et al., 2022). **Models**: For T1 MRI, we report two intrinsic atlases, one from the off-the-shelf model (Khader et al., 2022) and one from our age-conditioned model of Section 5, fine-tuned with the age condition disabled. Neither model is trained on IBSR18, Mindboggle101, or ABIDE.

	IBSR18				Mindboggle101				ABIDE			
	Dice↑	Cent.↓	$\|u\|$ ↓	SDlogJ↓	Dice↑	Cent.↓	$\|u\|$ ↓	SDlogJ↓	Dice↑	Cent.↓	$\|u\|$ ↓	SDlogJ↓
Baseline Methods												
voxel mean	0.723	2.65	3.40	0.084	0.507	2.85	3.65	0.086	0.665	2.59	3.21	0.076
ANTs	0.722	2.80	3.47	0.070	0.521	3.03	3.78	0.080	0.669	2.75	3.34	0.064
Deep learning methods												
VoxelMorph	0.732	2.24	3.07	0.071	0.527	2.59	3.49	0.086	0.674	2.16	2.92	0.065
Aladdin	0.730	2.82	3.56	0.082	0.512	3.23	4.01	0.093	0.664	2.67	3.32	0.075
MultiMorph $*$	**0.777**	1.70	2.83	0.077	**0.546**	2.40	3.47	0.094	**0.691**	1.67	2.77	0.072
Ours (off-the-shelf)	0.733	1.65	2.59	0.062	0.529	1.59	2.74	**0.065**	0.675	1.58	**2.43**	**0.055**
Ours	0.742	**1.52**	**2.53**	**0.060**	0.542	**1.41**	**2.72**	0.066	0.687	**1.53**	**2.43**	0.056

$*$: established atlases, without training on the same dataset.

(b) **On other domains**. **Metrics**: PCK@0.1 / @0.05↑ is the fraction of landmarks transferred through the template that land within 0.1 / 0.05 of the image size. IoU avg/mid denotes the mean / median of the part segmentation (e.g. seat, back, leg) by label transfer (Deng et al., 2021); 5 training data are deformed to its template, then deforming another data to validate the matches by nearest-neighbor voting. **Models**: generators trained by us on each domain's training split.

	CelebA (faces)	Mont. (x-ray)	KeypointNet (Chair)		KeypointNet (Car)	
	PCK@0.1/@0.05	PCK@0.1/@0.05	PCK@0.1/@0.05	IoU avg/mid	PCK@0.1/@0.05	IoU avg/mid
Baseline Methods						
pixel/voxel mean	0.613 / 0.396	0.675 / 0.455	0.252 / 0.051	0.603 / 0.605	0.739 / 0.333	0.615 / 0.626
ANTs	0.655 / 0.472	0.660 / 0.429	–	–	–	–
Deep Learning methods						
VoxelMorph	**0.694 / 0.489**	0.644 / 0.424	0.335 / 0.079	0.644 / 0.638	0.722 / 0.311	0.614 / 0.624
Aladdin	0.583 / 0.362	**0.727 / 0.514**	0.309 / 0.073	0.605 / 0.616	0.731 / 0.329	0.619 / 0.635
DIF-Net	–	–	0.448 / 0.114	0.668 / 0.656	**0.755** / 0.351	0.621 / 0.632
DIT	–	–	0.270 / 0.063	0.595 / 0.585	0.736 / 0.325	0.618 / 0.631
Ours	0.658 / 0.438	0.685 / 0.429	**0.465 / 0.117**	**0.696 / 0.679**	0.754 / **0.355**	**0.626 / 0.639**

VoxelMorph is first on faces (0.694 PCK@0.1) but fourth on chairs (0.335, 0.13 behind). Aladdin is first on chest X-ray (0.727) but last on faces (0.583, below the pixel mean). Our method is stronger than DIF-Net and DIT in PCK on both 3D shapes, and best on both PCK@0.05 and part-label IoU on 3D cars, second to DIF-Net by 0.001 on PCK@0.1.

5 On-Demand Atlas Families via Conditioning

If the intrinsic atlas is the central anatomy of the population, the model learned under condition y, then varying y should yield a whole family of intrinsic atlases from a single trained model. We test this with a continuous, clinically central conditioning variable: age. We fine-tune a brain-T1 generator with a continuous age condition on 899 T1 MRIs (IXI, OASIS-1; cognitively normal subjects, ages 18–94), and recover the atlas at a fixed age value with the same deterministic procedure, at guidance scale 1, with no per-age optimization. See supplementary for technical details.

Biological validity. An age-conditioned atlas family is only meaningful if it reproduces known age effects. One of the well-established effects in healthy aging is brain atrophy: the ventricles

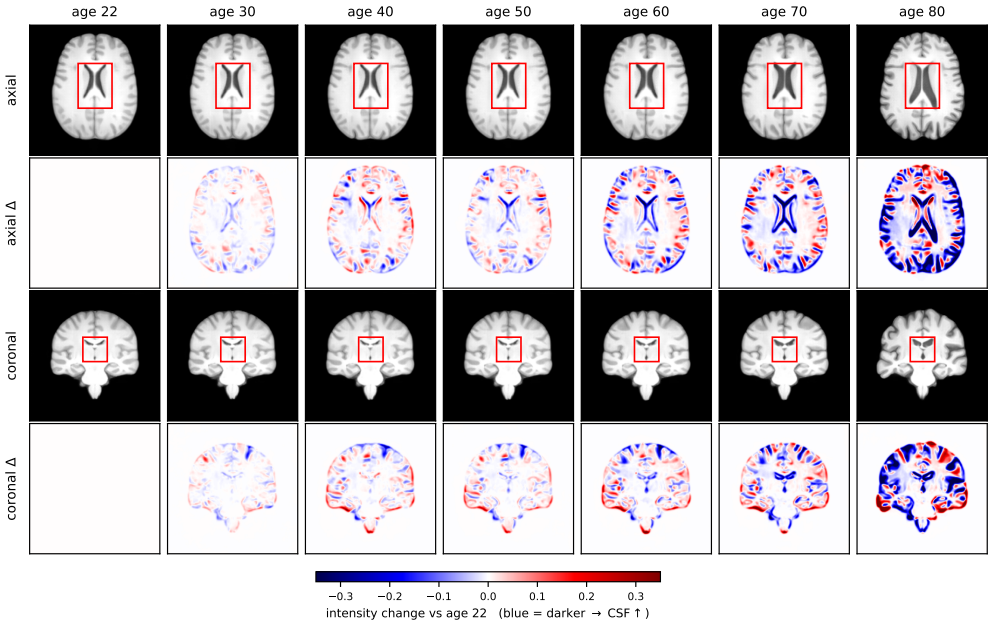
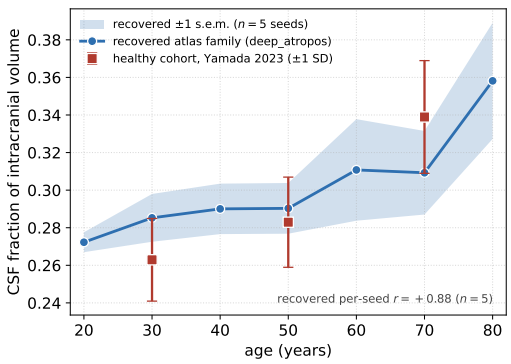
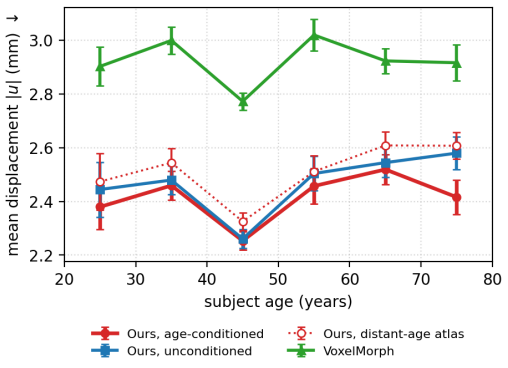


Figure 6: **The age-conditioned family, visually.** Columns are references recovered for ages 22–80 by varying only the age condition. Grayscale rows are the intrinsic atlases; rows beneath show the voxelwise difference from the youngest atlas on a shared diverging scale. Boxes (fixed across ages) frame the lateral ventricles; their enlargement is visible directly in grayscale.

enlarge, and the cerebrospinal-fluid (CSF) fraction of intracranial volume rises monotonically and increasingly steeply with age (Resnick et al., 2003; Fjell & Walhovd, 2010). We segment each recovered atlas with a contrast-robust deep-learning segmenter and track the CSF volume fraction, the canonical marker of brain aging. Figure 6 shows clear progressive ventricular enlargement during aging[1], as annotated in red squares in the figure. The recovered family reproduces the accelerating, monotonic CSF-expansion trend reported for cognitively normal cohorts (Yamada et al., 2023), consistently across five independent noise seeds (per-seed correlation $r{=}{+}0.88$, spread <0.2 percentage points at every age). Note that Yamada et al. (2023) report values from a commercial volumetry pipeline, whereas ours come from an open-source segmenter (Tustison et al., 2021a) with its own tissue boundaries, which produces a constant offset between the two curves. We therefore, in Figure 7a, compare the trend after removing a single constant shift.



(a) **The age-conditioned family reproduces the CSF-expansion signature of healthy aging.** Blue: CSF fraction vs. age (mean ± s.d., $n{=}5$ seeds). Red: independent reference values from Yamada et al. (2023), aligned by one disclosed constant shift.



(b) **Age-conditioned vs. unconditioned atlases**. Mean displacement $|u|$ on 78 subjects, grouped by decade. Age-matched atlases show a consistent gain.

[1]We show ages 22–80 to exclude the sparsely populated extremes of the 18–94 training range.

Registration utility. We test whether a conditioned atlas is a better registration target, as shown in Figure 7b, on the 78 IXI subjects of the cohort, aged 20–79 (the 45 OASIS-1 subjects are excluded). Compared to the non-age-conditioned atlas (blue line), the age-matched atlas (solid red line) requires less deformation for $53/78$ subjects and has the lower mean in every decade (-0.05 mm, $p{=}4{\times}10^{-5}$, $p < 0.001$). The VoxelMorph template trained on the same data lies a further 0.4–0.5 mm above it ($78/78$, $p{=}2{\times}10^{-14}$).

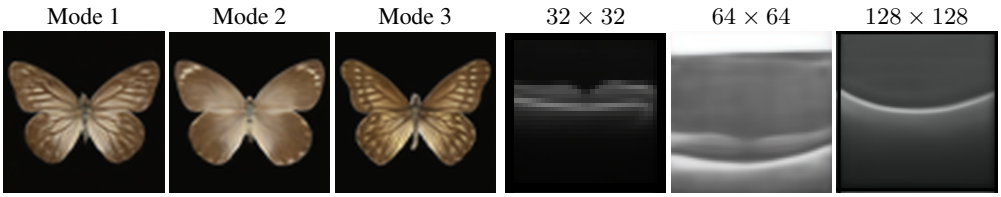
As a control, each subject is also registered to an atlas at a distant age (dotted red line, 25 for subjects aged 50 or older, 75 for younger ones). If conditioning carried no information, both would register equally well. Instead, the age-matched atlas gives a lower displacement error for 57 of 78 subjects, statistically significant ($p{=}5{\times}10^{-7}$, $p < 0.001$), and its mean displacement lies below the distant-age atlas in every decade. Conditioning on age produces atlases specific to that age, not merely plausible variants.

6 What Convergence Reveals About the Population

Our method always converges on the population's consensus, in whatever state that consensus exists. Figure 8 shows the three forms we observe when it is not a single sharp template.

Multiple templates. When a population contains distinct sub-groups, the initialization selects among them. Within one butterfly family of Pieridae, 60 seeds resolve to one of three similar pattern variants. Figure 8a shows three generated intrinsic atlases of the patterned sub-groups.

Scale-dependent template. A population can share structure at one modeling scale but not another: retinal OCT data contains a very simple structure, and its intrinsic atlases are coherent at each resolution but vary across resolutions, as shown in Figure 8b. Though every intrinsic atlas makes sense pixel-wise, the lower-resolution 32×32 atlas presents a more medically meaningful retinal structure with a clear fovea. Coherence is thus a pixel-space consensus, and the modeling scale decides which structures take part in it.



(a) Multiple population consensus.

(b) Scale-dependent population consensus



(c) **A remedy to restore the population consensus.** The same five FFHQ subjects, cropped to the face oval (top) and original (bottom), with the intrinsic atlas each population yields (right): a consensus structure gives a clear population atlas; a blurred image otherwise.

Figure 8: **What canonical convergence reveals. (a)** Multiple consensus: Pieridae seeds resolve to one of three pattern variants. **(b)** Scale-dependent consensus: OCT is coherent at $32{\times}32$ and degrades with resolution. **(c)** A remedy: the original frames (bottom) share no consensus and yield a blur; cropping the same subjects to the face oval (top) restores the consensus and a sharp atlas.

No template. When members vary in much of their content, the population consensus may not be a meaningful structure. As shown in Figure 8c, with a facial dataset, FFHQ, trained on original frames, where hair, background, and clothing vary, the convergence results in a featureless image. However, by manually removing those noises, the population-level consensus can be clearly reached.

7 Discussion

Classical construction computes a Fréchet mean, a statistical center defined by minimizing summed deformation under a chosen deformation model. The intrinsic atlas instead follows the generator's density, with no deformation model. By Tweedie's formula in Equation (2), each step of the ascent equals the mean displacement, $\mathbb{E}[x_0 - x \mid x]$, from the current expectation x to the population. The population itself, through the generator's density, defines the center. This is why the intrinsic atlas is central by construction, and the most central under registration in Table 1.

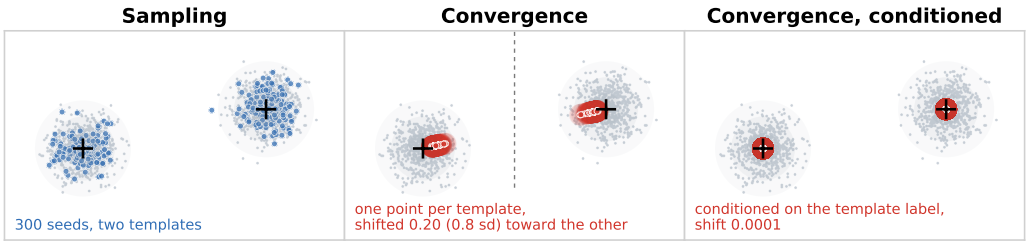


Figure 9: **Sampling versus convergence** on a population with two centers. *Left*: stochastic samples cover both. *Middle*: non-conditioned convergence finds two centers, but they pull on each other. *Right*: conditioned convergence finds two separated centers.

A structural bias, with a structural fix. If a population contains multiple centers, as shown in Figure 9, each recovered center is drawn toward the others, and conditioning restores their centering. The age family shows the same effect. The unconditioned atlas resembles the age-45 atlas most, close to the mean training age of 47, so it is already the atlas of the subjects in their 40s, and the corresponding conditioning shows a minimal gain. As shown in Figure 7b, the further a subject's age is from that mean, the more the conditioned atlas gains, least in the 40s and most in the 70s.

As the three Pieridae variants of Figure 8a show, seeds splitting into distinct images is itself a diagnostic that a population has several centers, with no label needed. The same principle behind mean-shift's use of multiple starts to find a density's modes (Comaniciu & Meer, 2002). A mode-detection tool, with a principled agreement threshold, is left for future work.

From a registration target to registration. The intrinsic atlas is a valid registration target, but the deformation onto it is still computed outside the generator. Prior congealing pipelines (Learned-Miller, 2006; Peebles et al., 2022; Ofri-Amar et al., 2023) require training to obtain correspondence. Diffusion-based works Tang et al. (2023); Luo et al. (2023); Kim et al. (2022) explored feature correspondence within diffusion models. We show that the template needs no training. If the correspondence does not either, congealing reduces entirely to inference on a generator built only to synthesize.

8 Conclusion

We set out from a hypothesis about the reverse process of a diffusion model: it combines a denoiser pulling toward the population and injected noise selecting an individual, so removing the noise should isolate what the population shares, its template in the sense of pattern theory. The recovered template, the *intrinsic atlas*, is an atlas in the practical sense: best or second-best on every dataset across brain MRI, faces, chest X-ray, and 3D shapes. Conditioning turns it into a family on demand, matched to any covariate the generator carries, and where a population holds several templates or none, the same process resolves into one image per template, or into a blurry image if none exists. Atlas construction, conventionally a dedicated optimization repeated for every population and covariate, is reframed as a byproduct of generative modeling: wherever a population coheres, and a generator has been trained on it, its atlas is already inside.

AI USE STATEMENT

We used generative AI tools to create scripts for preparing datasets and conducting experiments in non-brain MRI domains. The critical implementation and the evaluation of the T1 brain MRI domain originate from the authors. Additionally, we used generative AI tools to style figures and to polish sentences. We have reviewed all AI-assisted work. We verified the correctness of LLM-generated relevant literature in all phases. We reviewed the LLM-generated code and ensured its reproducibility. All drafted text was revised by the authors. We take responsibility for the final content of this work, including text, claims, or artifacts produced with the aid of generative AI.

REFERENCES

IXI Dataset (information extraction from images). https://brain-development.org/ixi-dataset/, 2005. Accessed: 2026-06-15.

S Mazdak Abulnaga, Andrew Hoopes, Neel Dey, Malte Hoffmann, Bruce Fischl, John Guttag, and Adrian Dalca. Multimorph: On-demand atlas construction. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 30906–30917, 2025.

B.B. Avants, C.L. Epstein, M. Grossman, and J.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. ISSN 1361-8415. doi: https://doi.org/10.1016/j.media.2007.06.004. URL https://www.sciencedirect.com/science/article/pii/S1361841507000606. Special Issue on The Third International Workshop on Biomedical Image Registration – WBIR 2006.

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.

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, February 2005. ISSN 1573-1405. doi: 10.1023/b:visi.0000043755.93987.aa. URL http://dx.doi.org/10.1023/B:VISI.0000043755.93987.aa.

BrainFacts/SfN. Internet brain segmentation repository. https://www.brainfacts.org/brainanatomy-and-function/anatomy/2012/mapping-the-brain, 2007. URL http://www.nitrc.org/projects/ibsr.

Dorin Comaniciu and Peter Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on pattern analysis and machine intelligence*, 24(5):603–619, 2002.

Adrian V. Dalca, Marianne Rakic, John Guttag, and Mert R. Sabuncu. *Learning conditional deformable templates with convolutional networks*. Curran Associates Inc., Red Hook, NY, USA, 2019.

Yu Deng, Jiaolong Yang, and Xin Tong. Deformed implicit field: Modeling 3d shapes with learned dense correspondence. In *IEEE Computer Vision and Pattern Recognition*, 2021.

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*, pp. 3929–3941, 2021.

Adriana Di Martino, Chao-Gan Yan, Qingyang Li, Erin Denio, Francisco X Castellanos, Kaat Alaerts, Jeffrey S Anderson, Michal Assaf, Susan Y Bookheimer, Mirella Dapretto, et al. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular psychiatry*, 19(6):659–667, 2014.

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*, pp. 20784–20793, 2022.

Bradley Efron. Tweedie's formula and selection bias. *Journal of the American Statistical Association*, 106(496):1602–1614, 2011.

Anders M Fjell and Kristine B Walhovd. Structural brain changes in aging: courses, causes and cognitive consequences. *Reviews in the Neurosciences*, 21(3):187–222, 2010.

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.

Ulf Grenander. A unified approach to pattern analysis. In *Advances in computers*, volume 10, pp. 175–216. Elsevier, 1970.

Ulf Grenander and Michael I Miller. Computational anatomy: An emerging discipline. *Quarterly of applied mathematics*, pp. 617–694, 1998.

Alessa Hering, Lasse Hansen, Tony CW Mok, Albert CS Chung, Hanna Siebert, Stephanie Häger, Annkristin Lange, Sven Kuckertz, Stefan Heldmann, Wei Shao, et al. Learn2reg: comprehensive multi-task medical image registration challenge, dataset and evaluation in the era of deep learning. *IEEE Transactions on Medical Imaging*, 42(3):697–712, 2022.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin (eds.), *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/4c5bcfec8584af0d967f1ab10179ca4b-Abstract.html.

Inbar Huberman-Spiegelglas, Vladimir Kulikov, and Tomer Michaeli. An edit friendly ddpm noise space: Inversion and manipulations. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12469–12478. IEEE, 2024.

Stefan Jaeger, Sema Candemir, Sameer Antani, Yì-Xiáng J Wáng, Pu-Xuan Lu, and George Thoma. Two public chest x-ray datasets for computer-aided screening of pulmonary diseases. *Quantitative imaging in medicine and surgery*, 4(6):475, 2014.

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

Firas Khader, Gustav Mueller-Franzes, Soroosh Tayebi Arasteh, Tianyu Han, Christoph Haarburger, Maximilian Schulze-Hagen, Philipp Schad, Sandy Engelhardt, Bettina Baessler, Sebastian Foersch, Johannes Stegmaier, Christiane Kuhl, Sven Nebelung, Jakob Nikolas Kather, and Daniel Truhn. Medical diffusion - denoising diffusion probabilistic models for 3d medical image generation, 2022. URL https://arxiv.org/abs/2211.03364.

Boah Kim, Inhwa Han, and Jong Chul Ye. Diffusemorph: Unsupervised deformable image registration using diffusion model. In *European conference on computer vision*, pp. 347–364. Springer, 2022.

Arno Klein and Jason Tourville. 101 labeled brain images and a consistent human cortical labeling protocol. *Frontiers in Neuroscience*, 6, 2012. ISSN 1662-4548. doi: 10.3389/fnins.2012.00171. URL http://dx.doi.org/10.3389/fnins.2012.00171.

Arno Klein, Jesper Andersson, Babak A. Ardekani, John Ashburner, Brian Avants, Ming-Chang Chiang, Gary E. Christensen, D. Louis Collins, James Gee, Pierre Hellier, Joo Hyun Song, Mark Jenkinson, Claude Lepage, Daniel Rueckert, Paul Thompson, Tom Vercauteren, Roger P. Woods, J. John Mann, and Ramin V. Parsey. Evaluation of 14 nonlinear deformation algorithms applied to human brain mri registration. *NeuroImage*, 46(3):786–802, July 2009. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2008.12.037. URL http://dx.doi.org/10.1016/j.neuroimage.2008.12.037.

Erik G Learned-Miller. Data driven image models through continuous joint alignment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(2):236–250, 2006.

Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *Proceedings of International Conference on Computer Vision (ICCV)*, December 2015.

Grace Luo, Lisa Dunlap, Dong Huk Park, Aleksander Holynski, and Trevor Darrell. Diffusion hyperfeatures: Searching through time and space for semantic correspondence. *Advances in Neural Information Processing Systems*, 36:47500–47510, 2023.

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. ISSN 1530-8898. doi: 10.1162/jocn.2007.19.9.1498. URL http://dx.doi.org/10.1162/jocn.2007.19.9.1498.

Chenlin Meng, Yutong He, Yang Song, Jiaming Song, Jiajun Wu, Jun-Yan Zhu, and Stefano Ermon. Sdedit: Guided image synthesis and editing with stochastic differential equations. *arXiv preprint arXiv:2108.01073*, 2021.

Dolev Ofri-Amar, Michal Geyer, Yoni Kasten, and Tali Dekel. Neural congealing: Aligning images to a joint semantic atlas. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 19403–19412. IEEE, 2023.

William Peebles, Jun-Yan Zhu, Richard Zhang, Antonio Torralba, Alexei A Efros, and Eli Shechtman. Gan-supervised dense visual alignment. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13460–13471. IEEE, 2022.

Marianne Rakic, Andrew Hoopes, S. Mazdak Abulnaga, Mert R. Sabuncu, John V. Guttag, and Adrian V. Dalca. Atlasmorph: Learning conditional deformable templates for brain MRI. *CoRR*, abs/2511.13609, 2025. doi: 10.48550/ARXIV.2511.13609. URL https://doi.org/10.48550/arXiv.2511.13609.

Amin Ranem, Camila González, Daniel Pinto dos Santos, Andreas M Bucher, Ahmed E Othman, and Anirban Mukhopadhyay. Continual atlas-based segmentation of prostate mri. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pp. 7563–7572, 2024.

Susan M Resnick, Dzung L Pham, Michael A Kraut, Alan B Zonderman, and Christos Davatzikos. Longitudinal magnetic resonance imaging studies of older adults: a shrinking brain. *The Journal of Neuroscience*, 23(8):3295–3301, 2003.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=PxTIG12RRHS.

Luming Tang, Menglin Jia, Qianqian Wang, Cheng Perng Phoo, and Bharath Hariharan. Emergent correspondence from image diffusion. *Advances in neural information processing systems*, 36: 1363–1389, 2023.

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, Michael A. Yassa, James R. Stone, James C. Gee, and Brian B. Avants. The antsx ecosystem for quantitative biological and medical imaging. *Scientific Reports*, 11(1), April 2021a. ISSN 2045-2322. doi: 10.1038/s41598-021-87564-6. URL http://dx.doi.org/10.1038/s41598-021-87564-6.

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, Michael A. Yassa, James R. Stone, James C. Gee, and Brian B. Avants. The ANTsX ecosystem for quantitative biological and medical imaging. *Scientific Reports*, 11(1):9068, April 2021b. ISSN 2045-2322. doi: 10.1038/s41598-021-87564-6. URL https://doi.org/10.1038/s41598-021-87564-6.

Zhenlin Xu and Marc Niethammer. Deepatlas: Joint semi-supervised learning of image registration and segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 420–429. Springer, 2019.

Shigeki Yamada, Tomohiro Otani, Satoshi Ii, Hiroto Kawano, Kazuhiko Nozaki, Shigeo Wada, Marie Oshima, and Yoshiyuki Watanabe. Aging-related volume changes in the brain and cerebrospinal fluid using artificial intelligence-automated segmentation. *European Radiology*, 33 (10):7099–7112, April 2023. ISSN 1432-1084. doi: 10.1007/s00330-023-09632-x. URL http://dx.doi.org/10.1007/s00330-023-09632-x.

Jiancheng Yang, Udaranga Wickramasinghe, Bingbing Ni, and Pascal Fua. Implicitatlas: Learning deformable shape templates in medical imaging. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022*, pp. 15840–15850. IEEE, 2022. doi: 10.1109/CVPR52688.2022.01540. URL https://doi.org/10.1109/CVPR52688.2022.01540.

Yang You, Yujing Lou, Chengkun Li, Zhoujun Cheng, Liangwei Li, Lizhuang Ma, Cewu Lu, and Weiming Wang. Keypointnet: A large-scale 3d keypoint dataset aggregated from numerous human annotations. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13644–13653. IEEE, 2020.

Zerong Zheng, Tao Yu, Qionghai Dai, and Yebin Liu. Deep implicit templates for 3d shape representation. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1429–1439. IEEE, 2021.