Generating Novel Brain Morphology by Deforming Learned Templates

Alan Q. Wang[1], Fangrui Huang[1], Bailey Trang[1], Wei Peng[1], Mohammad Abbasi[1], Kilian Pohl[1], Mert R. Sabuncu[2], and Ehsan Adeli[1]

[1] Stanford University, Stanford, CA 94305, USA
[2] Cornell Tech and Weill Cornell Medicine, New York, NY 10044, USA
alanqw@stanford.edu

Abstract. Designing generative models for 3D structural brain MRI that synthesize morphologically-plausible and attribute-specific (e.g., age, sex, disease state) samples is an active area of research. Existing approaches based on frameworks like GANs or diffusion models synthesize the image directly, which may limit their ability to capture intricate morphological details. In this work, we propose a 3D brain MRI generation method based on state-of-the-art latent diffusion models (LDMs), called MorphLDM, that generates novel images by applying synthesized deformation fields to a learned template. Instead of using a reconstruction-based autoencoder (as in a typical LDM), our encoder outputs a latent embedding derived from both an image and a learned template that is itself the output of a template decoder; this latent is passed to a deformation field decoder, whose output is applied to the learned template. A registration loss is minimized between the original image and the deformed template with respect to the encoder and both decoders. Empirically, our approach outperforms generative baselines on metrics spanning image diversity, adherence with respect to input conditions, and voxel-based morphometry.
Our code is available at `https://github.com/alanqrwang/morphldm`.

Keywords: MRI Generation · Morphology · Deformable Templates.

1 Introduction

Morphology is an important characteristic of neuroimaging, known to be related to attributes and/or phenotypes such as age, sex, and disease state [20, 31]. A growing literature aims to design generative models for structural brain magnetic resonance imaging (MRI), which synthesize realistic morphology and are specific to conditioned attributes such as age and sex [21, 23]. These models offer the potential to augment limited training data [24], generate counterfactuals [15], and increase interpretability [30].

In computer vision, generative models have found success in natural image domains, driven largely by GANs [8, 10, 35] and diffusion models [13, 26]. These

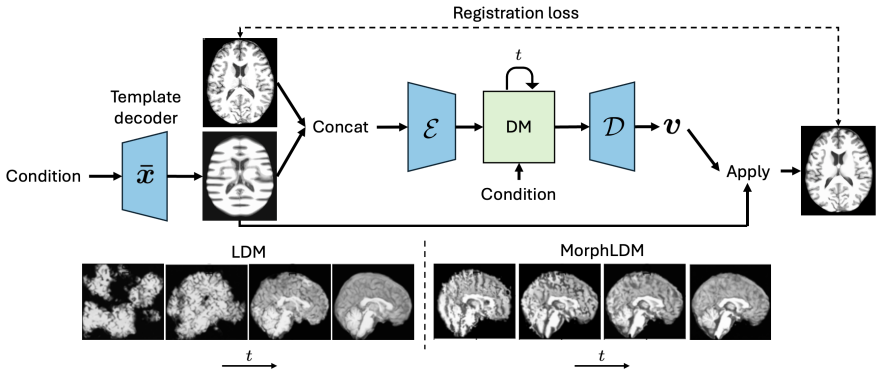

Fig. 1. (top) Graphical depiction of MorphLDM. The encoder takes in an image and a template that is the output of a template decoder, concatenated along the channel dimension. The decoder outputs a deformation field $\boldsymbol{v}$ that is applied to the template. Encoder-decoder networks in blue are trained in the first stage on a registration loss (similar to LDM). The diffusion model (DM) in green is trained on the learned latents in the second stage. (bottom) Visualization of denoising steps $t \in [T]$ of LDM (left) and MorphLDM (right). While LDM's decoder maps from a synthesized latent to an image directly, MorphLDM's decoder maps to a synthesized deformation field, which is subsequently applied to the learned template.

models work by learning the data distribution $p(\boldsymbol{x})$ from which the data is assumed to be sampled. Although highly general and capable of capturing the variety inherent in naturalistic images, this approach applied to neuroimaging may fail to capture intricacies and subtleties crucial to morphology, especially in the context of high-dimensional 3D volumetric data and data-limited scenarios.

Specific to neuroimaging, previous work has shown that much of the variability between individual anatomies can be captured by geometric and topological changes [4, 5]. This motivates a line of work that, instead of directly synthesizing images, synthesizes dense deformation fields that are applied to a template image. However, in addition to using less powerful generative models like GANs, prior work either precomputes deformation fields using an expensive registration method [3, 37] or fixes the template, thus biasing and limiting the model's ability to adequately capture the image distribution [3, 17, 36].

In this work, we propose a 3D brain MRI generation method based on state-of-the-art latent diffusion models (LDMs), called MorphLDM, that generates novel images by applying synthesized deformation fields to a *learned* template. MorphLDM differs from LDMs in the design of the encoder/decoder. First, a learned template is outputted by a template decoder, optionally conditioned on image-level attributes. Then, an encoder takes in both an image and the template and outputs a latent embedding; this latent is passed to a deformation field decoder, whose output deformation field is applied to the template. Finally, a registration loss is minimized between the original image and the deformed tem-

plate with respect to the encoder and both decoders. Subsequently, a diffusion model is trained on these learned latent embeddings.

To synthesize an image, MorphLDM generates a novel latent in the same way as standard LDMs. The decoder maps this latent to its corresponding deformation field, which is subsequently applied to the learned template. We find that synthetic samples generated by our approach not only outperform powerful baselines on standard image generation metrics like FID, but also better capture input conditions known to be related to morphology, including age and sex. Additionally, we perform a voxel-based morphometry [33] analysis and find that our models generate more realistic structures in terms of regional volume compared to baselines.

To the best of our knowledge, MorphLDM is the first MRI diffusion model that synthesizes deformation fields applied to a learned template. Our approach is efficient and data-driven in the sense that both the deformation fields and the template are learned simultaneously, without requiring precomputation of deformation fields or making a specific choice for the template.

2 Background

2.1 Brain MRI Generation

Generative adversarial networks (GANs) and their extensions have been widely explored in the context of brain MRIs, including using Wasserstein GANs [11], incorporating variational autoencoders (VAEs) and code discriminators [18], and leveraging segmentations [8]. Autoregressive (AR) approaches generate samples in an autoregressive fashion and require discrete tokenization of images, typically done with a vector-quantized (VQ)-VAE [7, 28].

More recently, latent diffusion models (LDMs) have become the state-of-the-art generative model for brain MRIs [21, 23], and are trained in two steps. In the first stage, an autoencoder $\mathcal{D}(\mathcal{E}(\boldsymbol{x})) = \mathcal{D}(\boldsymbol{z})$ is trained on a reconstruction loss $\mathcal{L}_{sim}$ to learn a latent space. In the second stage, a diffusion model is trained on the latent space defined by $\boldsymbol{z} = \mathcal{E}(\boldsymbol{x})$; specifically, a network ϵ_ω is trained to perform time-conditioned denoising of the latent embeddings:

$$\underset{\omega}{\operatorname{argmin}}\, \mathbb{E}_{\mathcal{E}(\boldsymbol{x}),\epsilon\sim\mathcal{N}(0,1),t\in[T]}||\epsilon - \epsilon_\omega(\boldsymbol{z}_t, t, \boldsymbol{c})||_2^2, \tag{1}$$

where $\mathbb{E}$ is expected value, $\mathcal{N}(0, I)$ is the standard normal distribution, and $\boldsymbol{c}$ is any conditioning information associated with the image.[3]

2.2 Templates

Deformable templates, or atlases, are common in computational anatomy. Research is dedicated to constructing unbiased templates such that the deformation

[3] Typically, ϵ_ω is implemented as a time-conditional UNet [26], where the conditioning information $\boldsymbol{c}$ is included via cross-attention with the intermediate feature maps.

fields from these templates to individual images can be analyzed to understand population variability [2, 4, 9]. More recently, template learning approaches use neural networks to learn templates that are data-driven and attribute-specific [5, 6, 27]. These approaches aim to build decoders that learn templates conditioned on specified attributes. In this work, we take inspiration from template learning approaches, where our goal is to learn a universal template concurrently with a deformation field synthesizer to generate novel brain images.

3 Methods

Let $\boldsymbol{x} \in \mathbb{R}^{C \times L \times W \times H}$ denote a 3D volume sampled from a distribution $p(\boldsymbol{x})$. For MRIs, $\boldsymbol{x}$ is a single-channel intensity image where $C = 1$. Let $\boldsymbol{z} \sim p(\boldsymbol{z})$ denote a sample from some prior distribution (usually a standard normal) and let $\boldsymbol{c} \in \mathcal{C}$ denote (optional) conditioning information. Typical generative approaches aim to model $p(\boldsymbol{x})$ directly via a generator $\mathcal{G}$ that takes in a latent sample along with the condition: $\hat{\boldsymbol{x}} = \mathcal{G}(\boldsymbol{z}, \boldsymbol{c})$.

We assume the existence of a template $\bar{\boldsymbol{x}} \in \mathbb{R}^{C \times L \times W \times H}$ such that any $\boldsymbol{x} \sim p(\boldsymbol{x})$ is related to $\bar{\boldsymbol{x}}$ via applying a deformation field $\boldsymbol{v} \in \mathbb{R}^{3 \times L \times W \times H}$: $\boldsymbol{x} = \boldsymbol{v} \circ \bar{\boldsymbol{x}}$. Under the template assumption, we can recast modeling $p(\boldsymbol{x})$ as modeling $p(\boldsymbol{v})$, the distribution of deformation fields induced by $\bar{\boldsymbol{x}}$. Then, the output of the generator is a novel deformation field $\hat{\boldsymbol{v}} = \mathcal{G}(\boldsymbol{z}, \bar{\boldsymbol{x}}, \boldsymbol{c})$, and the novel image derived from this deformation field is $\hat{\boldsymbol{x}} = \hat{\boldsymbol{v}} \circ \bar{\boldsymbol{x}}$.[4]

To prevent any bias due to a specific choice of $\bar{\boldsymbol{x}}$, we learn $\bar{\boldsymbol{x}}$ concurrently with $\mathcal{G}$ by optimizing a separate template decoder $\bar{\boldsymbol{x}} : \mathcal{C} \rightarrow \mathbb{R}^{C \times L \times W \times H}$, which outputs a deterministic, universal template on which all synthesized deformation fields are applied. In the simplest case, $\bar{\boldsymbol{x}}$ can be made unconditional by setting the input vector to be learnable. We also experiment with adding $\boldsymbol{c}$ as an input to the template decoder to learn conditional templates. This enables the learned templates to capture intensity or attribute-specific structural detail that can improve overall generation [5, 6].

3.1 MorphLDM

In this work, we apply our approach to state-of-the-art LDMs, where the second stage diffusion training remains unchanged (Eq. (1)) and the main changes are in the design and training of the autoencoder. Fig. 1 gives a graphical depiction. Let $\mathcal{E}_\theta$ and $\mathcal{D}_\phi$ denote the image encoder and decoder parameterized by θ and ϕ, respectively. The input to the autoencoder is $\boldsymbol{x}$ and $\bar{\boldsymbol{x}}$ concatenated along the channel dimension. The encoder output $\boldsymbol{z}$ is passed to the decoder, which outputs a deformation field $\boldsymbol{v}$ that is applied to the template $\bar{\boldsymbol{x}}$ via a differentiable grid-sampler. Thus, the objective is

$$\underset{\bar{\boldsymbol{x}}, \theta, \phi}{\operatorname{argmin}} \, \mathbb{E}_{\boldsymbol{x}} \left[\mathcal{L}_{sim}(\mathcal{D}_\phi(\mathcal{E}_\theta(\boldsymbol{x}, \bar{\boldsymbol{x}})) \circ \bar{\boldsymbol{x}}, \boldsymbol{x}) + \mathcal{R}(\boldsymbol{v}) \right], \tag{2}$$

[4] The template assumption is common and well-motivated in the literature [3–5], but notably may be violated when $\boldsymbol{x}$ exhibits lesions or large structural abnormalities.

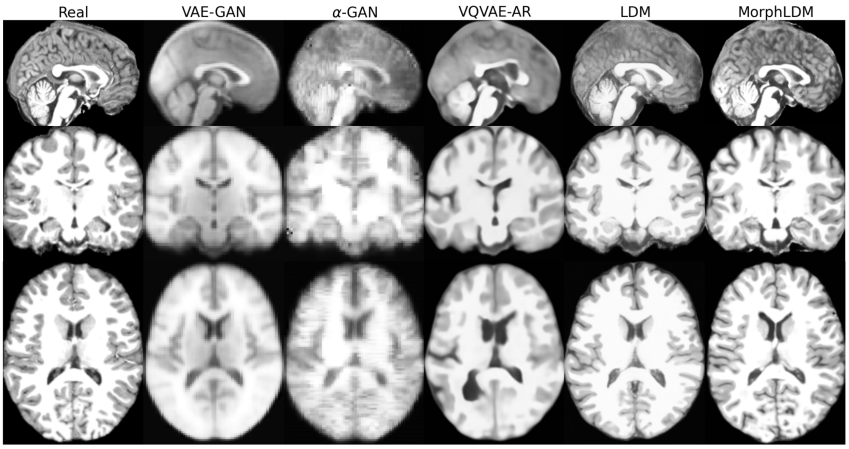

Fig. 2. Representative real and synthetic samples for all models shown in the three planes. Age=10, sex="female". We observe that MorphLDM has a higher degree of morphological detail compared to baselines, especially in folding patterns in cortical regions.

where $\mathcal{R}(\boldsymbol{v})$ denotes a regularization term on the predicted deformation field. In this work, we penalize the magnitude and spatial gradient of the displacement field. Letting $\boldsymbol{u_v}$ denote the spatial displacement for $\boldsymbol{v} = Id + \boldsymbol{u_v}$, where Id is the identity transformation: $\mathcal{R}(\boldsymbol{v}) = \alpha||\boldsymbol{u_v}||_2 + \beta||\nabla \boldsymbol{u_v}||_2$. Following prior work [23, 26], we also impose a slight KL penalty towards a standard normal on the latent space to prevent arbitrarily high-variance latent spaces.[5].

4 Experiments

Dataset. We train all models on a large dataset of publicly-available T1-weighted brain MRIs.[6] For all datasets, we restrict to healthy controls when applicable. In total, we have 27,066 3D volumes, of which we reserve 21,051 for training and the rest for validation. Each volume is of resolution 160x192x176 and skullstripped and registered to MNI space. Age and sex metadata is available for all images. Since our dataset is skewed toward younger ages, we uniformly sample ages binned across decades during training.

For generative model evaluation, we generate 1000 samples with conditions linearly spaced between the ages of 5 to 100 and sexes are evenly split between

[5] Many works in this space utilize diffeomorphic constraints [2, 5, 24]. We found experimentally that this constraint was too strong and led to artifacts in the synthesized samples.

[6] Data are from the Alzheimer's Disease Neuroimaging Initiative (ADNI) [22], Adolescent Brain Cognitive Development (ABCD) Study [16], Human Connectome Project (HCP) [29], and Parkinson's Progression Markers Initiative (PPMI) [25].

Table 1. Image quality metrics. ↑: higher is better, ↓: lower is better

	Sex Acc ↑	Age MAE ↓	FID ↓	MS-SSIM ↓
Real	0.88	4.63	–	0.74
VAE-GAN	0.55	35.53	298.38	0.81
α-GAN [18]	0.51	33.30	313.58	0.90
VQVAE-AR [28]	0.62	29.62	290.01	0.80
LDM [23]	0.78	12.95	273.71	0.77
LDM^{c}	0.79	10.37	282.52	0.79
MorphLDM (ours)	0.84	5.59	213.70	**0.73**
MorphLDM^{c} (ours)	**0.85**	**4.93**	**202.49**	**0.73**

male and female. For real samples, we collect 1000 validation samples by collecting 500 female validation samples closest in age to the synthetic samples (without replacement), and similarly for male samples.

Architectural and Training Details. For all LDM-based models, the encoder $\mathcal{E}$ is composed of convolutional blocks with 3 levels of downsampling; the number of latent channels is 8. For LDM, $\mathcal{D}$ outputs a 1-channel reconstruction. For MorphLDM, it outputs a 3-channel output for a 3D deformation field. The diffusion UNet ϵ_ω has intermediate channel sizes of [384, 512, 512], and cross-attention is applied with $\boldsymbol{c}$ at the latter two levels. The architecture of the template decoder $\bar{\boldsymbol{x}}$ is similar to the decoder for the deformation field, except that it takes as input a vector with a size equal to the number of conditions $\boldsymbol{c}$ and outputs a template $\bar{\boldsymbol{x}} \in \mathbb{R}^{C \times L \times W \times H}$.

For training, we use the same reconstruction/similarity loss $\mathcal{L}_{sim}$ for all LDM-based models, which is composed of an L1 loss and an adversarial loss using a patch-based discriminator [14] with a weight of 0.005. A loss coefficient of $1e-7$ is used for the KL penalty. For MorphLDM, we set $\alpha = 5$ and $\beta = 1$. All training and evaluation is done on an NVIDIA H100 GPU.

Baselines. We compare against two GAN-based models, VAE-GAN and α-GAN [18], and an autoregressive model (AR) trained on discrete tokens learned through a VQVAE [28]. We also compare against an LDM model, which minimizes a reconstruction loss $\mathcal{L}_{sim}$ during autoencoder training. All architectural and training details are kept as identical as possible between LDM and MorphLDM variants.

All LDM-based models have conditioning information $\boldsymbol{c}$ injected in the diffusion UNet ϵ_ω via cross-attention with the intermediate feature maps. We also experiment with including $\boldsymbol{c}$ in the autoencoder by concatenating age and sex as additional channels of $\boldsymbol{z}$ repeated across the spatial grid. For MorphLDM, $\boldsymbol{c}$ is also passed as input to the template decoder $\bar{\boldsymbol{x}}$. Throughout this subsection, LDM-based models which have conditioning information injected in the encoder/decoder are denoted with a superscript ^{c}.

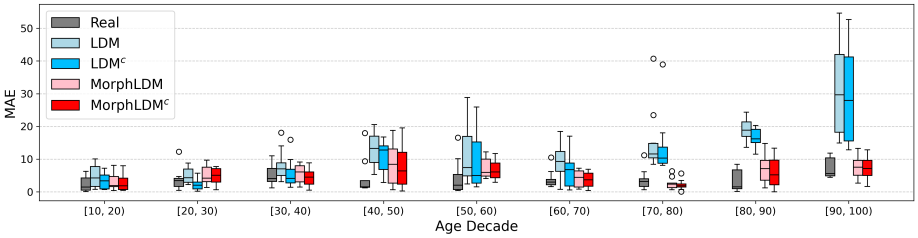

Fig. 3. Prediction error of synthetic samples on a pretrained age predictor across age decades. "Real" denotes prediction error on real (i.e. validation) data. Generally, we find that MorphLDM models exhibit lower MAE values than the LDM model across all age ranges, especially at older ages, where less training data is available. MorphLDM^c generally outperforms MorphLDM.

Metrics. We report the Frechet inception distance (FID) [12] using pretrained ImageNet weights, which has been shown to more closely align with expert judgment [34]. To quantify diversity, we report the multi-scale structural similarity measure (MS-SSIM) [23, 32], which is computed by averaging over 1000 MS-SSIM values for 1000 pairs of images. Lower is better since a higher MS-SSIM indicates a higher degree of image similarity and therefore, lower diversity.

To measure how well synthetic samples capture age and sex information, we pass the samples through pretrained models for age and sex and quantify the prediction error. We train a CNN-based age regressor and sex classifier on the same (real) data used for training the generative models.[7]

We perform a voxel-based morphometry analysis to evaluate synthetic samples with respect to volumes of brain regions [33]. To obtain segmentations for both real and synthetic data, we use SynthSeg [1], a widely-used tool for segmenting brain MRI volumes.

4.1 Image Generation

Fig. 2 depicts a representative sample across three views for all models. We observe that LDM-based samples exhibit sharper images, with GAN-based models producing blurry results as well as less diversity across samples. We observe that MorphLDM has a higher degree of morphological detail compared to LDM, especially in folding patterns in cortical regions. This may be attributed to the model's ability to focus on morphology instead of other aspects of the image.[8]

[7] The architectures for the age and sex predictors are CNNs with 4 downsampling levels and 2 convolutional blocks (conv, norm, and ReLU) per level. The final layer is a channel-wise average pooling followed by a fully-connected layer down to a scalar output. The age regressor is trained to minimize MSE loss and the sex classifier is trained to minimize binary cross entropy.

[8] We refer the reader to the Supplementary material for a visualization of the learned templates for MorphLDM^c.

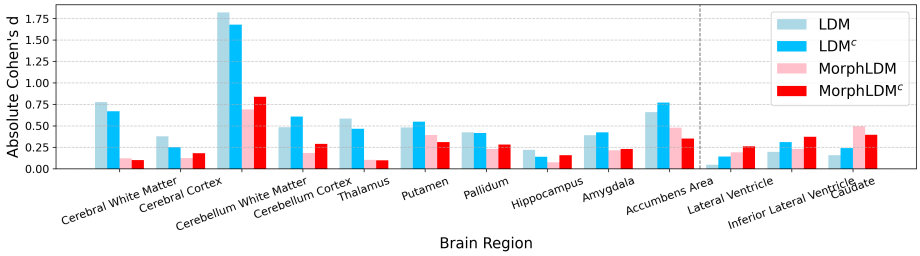

Fig. 4. Absolute Cohen's d of regional volumes between real and synthetic samples. Lower is better. MorphLDM variants exhibit lower values for 10 out of 13 regions.

Table 1 summarizes the results on various image generation metrics. We find that both MorphLDM variants better capture age and sex information, as evidenced by higher sex accuracy and lower age MAE. MorphLDM^c generally outperforms MorphLDM, which can be attributed to its increased expressivity in being able to generate age-conditioned templates. MorphLDM variants also outperform LDM variants on FID and MS-SSIM. We attribute this to MorphLDM's ability to capture more detailed and varied morphology compared to LDM.

Fig. 3 plots the age MAE across age decades for all real and synthetic samples. The improvement of MorphLDM over baseline LDM becomes more pronounced as age increases. Given that our training dataset is skewed toward brain MRIs of younger subjects (5-20 years), we hypothesize that MorphLDM is more data-efficient due to restricting its modeling to morphology.

4.2 Effect Size of Regional Volumes

To quantify the morphological realism of synthetic images, we quantify the effect size of regional volumes between populations of synthetic and real brains [21, 19] using the absolute Cohen's d.[9] Fig. 4 shows a bar graph of absolute Cohen's d values for different models, across all brain regions that are segmented by SynthSeg. We find that MorphLDM models exhibit lower Cohen's d values for 10 out of the 13 regions (left of the dotted line). This indicates a broadly closer degree of anatomical similarity with real samples in terms of regional volumes.

5 Conclusion

We propose MorphLDM, a diffusion model for structural brain MRI generation that synthesizes novel displacement fields applied to a learned, universal template. The "autoencoder" of MorphLDM takes as input an image and learnable template and outputs a dense deformation field; this deformation field is applied

[9] Absolute Cohen's d is defined as the absolute difference between the two population means divided by the pooled standard deviation: $d = \frac{|\bar{x}_1 - \bar{x}_2|}{s}$.

to the learnable template, and a registration loss is minimized between the original image and the deformed template. Empirically, our approach outperforms baselines with respect to standard image quality metrics, faithfulness to specified attribute conditions, and voxel-based morphometry.

Acknowledgments. This work was supported in part by National Institutes of Health grant AG089169. This study was also supported by the Stanford School of Medicine Department of Psychiatry and Behavioral Sciences Jaswa Innovator Award, the Stanford Institute for Human-Centered AI Postdoctoral Fellowship, and the Stanford HAI Hoffman-Yee Award.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Billot, B., Greve, D.N., Puonti, O., Thielscher, A., Van Leemput, K., Fischl, B., Dalca, A.V., Iglesias, J.E.: Synthseg: Segmentation of brain mri scans of any contrast and resolution without retraining. Medical Image Analysis **86**, 102789 (2023)
2. Blaiotta, C., Freund, P., Cardoso, M.J., Ashburner, J.: Generative diffeomorphic modelling of large mri data sets for probabilistic template construction. NeuroImage **166**, 117–134 (2018)
3. Chong, C.K., Ho, E.T.W.: Synthesis of 3d mri brain images with shape and texture generative adversarial deep neural networks. IEEE Access **9**, 64747–64760 (2021). https://doi.org/10.1109/ACCESS.2021.3075608
4. Christensen, G., Joshi, S., Miller, M.: Volumetric transformation of brain anatomy. IEEE Transactions on Medical Imaging **16**(6), 864–877 (1997)
5. Dalca, A., Rakic, M., Guttag, J., Sabuncu, M.: Learning conditional deformable templates with convolutional networks. In: NeurIPS. vol. 32 (2019)
6. Dey, N., Ren, M., Dalca, A.V., Gerig, G.: Generative adversarial registration for improved conditional deformable templates (2022)
7. Esser, P., Rombach, R., Ommer, B.: Taming transformers for high-resolution image synthesis (2021)
8. Fernandez, V., Pinaya, W.H.L., Borges, P., Graham, M.S., Tudosiu, P.D., Vercauteren, T., Cardoso, M.J.: Generating multi-pathological and multi-modal images and labels for brain mri. Medical Image Analysis **97**, 103278 (2024)
9. Ganser, K.A., Dickhaus, H., Metzner, R., Wirtz, C.R.: A deformable digital brain atlas system according to talairach and tournoux. Medical Image Analysis **8**(1), 3–22 (2004)
10. Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial networks (2014)
11. Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A.: Improved training of wasserstein gans (2017)
12. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: Gans trained by a two time-scale update rule converge to a local nash equilibrium (2018)
13. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models (2020)

14. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks (2018)
15. Jeanneret, G., Simon, L., Jurie, F.: Diffusion models for counterfactual explanations (2022)
16. Karcher, N.R., Barch, D.M.: The abcd study: understanding the development of risk for mental and physical health outcomes. Neuropsychopharmacology **46**(1), 131–142 (2021)
17. Kim, B., Ye, J.C.: Diffusion deformable model for 4d temporal medical image generation (2022), https://arxiv.org/abs/2206.13295
18. Kwon, G., Han, C., Kim, D.s.: Generation of 3d brain mri using auto-encoding generative adversarial networks. In: MICCAI 2019. pp. 118–126 (2019)
19. Lakens, D.: Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and anovas. Frontiers in Psychology **4** (2013)
20. Madan, C.R.: Advances in studying brain morphology: The benefits of open-access data. Frontiers in Human Neuroscience **11** (2017)
21. Peng, W., Bosschieter, T., Ouyang, J., Paul, R., Sullivan, E.V., Pfefferbaum, A., Adeli, E., Zhao, Q., Pohl, K.M.: Metadata-conditioned generative models to synthesize anatomically-plausible 3d brain mris. Medical Image Analysis **98**, 103325 (2024)
22. Petersen, R.C., Aisen, P.S., Beckett, L.A., Donohue, M.C., Gamst, A.C., Harvey, D.J., Jack, Clifford R., J., Jagust, W.J., Shaw, L.M., Toga, A.W., Trojanowski, J.Q., Weiner, M.W.: Alzheimer's disease neuroimaging initiative (adni): clinical characterization. Neurology **74**(3), 201–209 (2010)
23. Pinaya, W.H.L., Tudosiu, P.D., Dafflon, J., da Costa, P.F., Fernandez, V., Nachev, P., Ourselin, S., Cardoso, M.J.: Brain imaging generation with latent diffusion models (2022)
24. Pombo, G., Gray, R., Cardoso, M.J., Ourselin, S., Rees, G., Ashburner, J., Nachev, P.: Equitable modelling of brain imaging by counterfactual augmentation with morphologically constrained 3d deep generative models. Medical Image Analysis **84**, 102723 (2023)
25. Pulliam, E., Singleton, A.B.: The parkinson's progression markers initiative (ppmi): Study design and protocol. Movement Disorders **26**(9), 1453–1460 (2011)
26. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models (2022)
27. Starck, S., Sideri-Lampretsa, V., Kainz, B., Menten, M., Mueller, T., Rueckert, D.: Diff-def: Diffusion-generated deformation fields for conditional atlases (2024), https://arxiv.org/abs/2403.16776
28. Tudosiu, P.D., Pinaya, W.H.L., Graham, M.S., Borges, P., Fernandez, V., Yang, D., Appleyard, J., Novati, G., Mehra, D., Vella, M., Nachev, P., Ourselin, S., Cardoso, J.: Morphology-preserving autoregressive 3d generative modelling of the brain. In: Simulation and Synthesis in Medical Imaging: 7th International Workshop, SASHIMI 2022, Held in Conjunction with MICCAI 2022, Singapore, September 18, 2022, Proceedings. p. 66–78 (2022)
29. Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E., Yacoub, E., Ugurbil, K.: The wu-minn human connectome project: An overview. NeuroImage **80**, 62–79 (2013), mapping the Connectome
30. Wang, A.Q., Karaman, B.K., Kim, H., Rosenthal, J., Saluja, R., Young, S.I., Sabuncu, M.R.: A framework for interpretability in machine learning for medical imaging. IEEE Access **12**, 53277–53292 (2024)

31. Wang, Y., Leiberg, K., Ludwig, T., Little, B., Necus, J.H., Winston, G., Vos, S.B., de Tisi, J., Duncan, J.S., Taylor, P.N., Mota, B.: Independent components of human brain morphology. NeuroImage **226**, 117546 (2021)
32. Wang, Z., Simoncelli, E., Bovik, A.: Multiscale structural similarity for image quality assessment. In: Asilomar Conference. vol. 2, pp. 1398–1402 (2003)
33. Whitwell, J.L.: Voxel-based morphometry: An automated technique for assessing structural changes in the brain. Journal of Neuroscience **29**(31), 9661–9664 (2009)
34. Woodland, M., Castelo, A., Al Taie, M., Albuquerque Marques Silva, J., Eltaher, M., Mohn, F., Shieh, A., Kundu, S., Yung, J.P., Patel, A.B., Brock, K.K.: Feature extraction for generative medical imaging evaluation: New evidence against an evolving trend. In: MICCAI 2024. pp. 87–97 (2024)
35. Yi, X., Walia, E., Babyn, P.: Generative adversarial network in medical imaging: A review. Medical Image Analysis **58**, 101552 (2019)
36. Zheng, J.Q., Mo, Y., Sun, Y., Li, J., Wu, F., Wang, Z., Vincent, T., Papież, B.W.: Deformation-recovery diffusion model (drdm): Instance deformation for image manipulation and synthesis (2024), `https://arxiv.org/abs/2407.07295`
37. Zhuo, Y., Shen, Y.: Diffusereg: Denoising diffusion model for obtaining deformation fields in unsupervised deformable image registration (2024), `https://arxiv.org/abs/2410.05234`