Mitigating analytical variability in fMRI results with style transfer

Elodie Germani, Camille Maumet, Elisa Fromont

Mitigating analytical variability in fMRI results with style transfer

Elodie Germani[1], Camille Maumet[2], and Elisa Fromont[1]

[1] Univ Rennes, Inria, CNRS, Inserm, Rennes, France
[2] Univ Rennes, IUF, Inria, CNRS, Rennes, France
{elodie.germani},{elisa.fromont}@irisa.fr

Abstract. We propose a novel approach to improve the reproducibility of neuroimaging results by converting statistic maps across different functional MRI pipelines. We make the assumption that pipelines can be considered as a style component of data and propose to use different generative models, among which, Diffusion Models (DM) to convert data between pipelines. We design a new DM-based unsupervised multi-domain image-to-image transition framework and constrain the generation of 3D fMRI statistic maps using the latent space of an auxiliary classifier that distinguishes statistic maps from different pipelines. We extend traditional sampling techniques used in DM to improve the transition performance. Our experiments demonstrate that our proposed methods are successful: pipelines can indeed be transferred, providing an important source of data augmentation for future medical studies.

Keywords: Neuroimaging · Analytical variability · Deep learning · Style transfer · Diffusion models · Generative models.

1 Introduction

For the past ten years, scientific research faced a "reproducibility crisis", prompting the adoption of new research practices, in particular to increase sample sizes. Data sharing platforms [18,8] were developed to facilitate the re-use of raw data but also of derived data through meta- and mega-analyses [5]. In functional Magnetic Resonance Imaging (fMRI), due to the high flexibility of the analytical pipelines [2], derived data shared on public databases often come from different workflows, also known as 'pipelines'. However, different pipelines can lead to different results [1] and combining results from different pipelines in mega-analyses can lead to a higher risk of false positive findings [26]. To benefit from these large amount of derived data available, it is necessary to find a way to mitigate the effect of analytical variability.

In computer vision, recent advances gave rise to performing deep generative models such as Generative Adversarial Networks (GANs) [7] and Denoising Diffusion Probabilistic Models (DDPMs) [9]. These models produce high quality results for generating new images from a known distribution but also in the task of image-to-image transition [11,28], i.e. converting data from a source domain to

a target domain. Unsupervised frameworks [32,30,15] do not necessitate pairs of data in different domains for their training as they use constraints like cycle consistency [32] and shared latent space assumption [15,30]. They provide a good opportunity to benefit from large unlabeled databases to learn complex mapping without any ground-truth target data. By conditioning on domain-specific features (e.g. class vector, latent space of auxiliary classifiers, etc.) instead of full target images, unsupervised frameworks also extend to multi-domain transitions [4,3,10] to learn transfer between multiple domains in a single model.

Recently, DDPMs have achieved state-of-the-art performance in synthesizing natural images, overpassing GANs by producing complex and diverse images [20], while reducing the risk of modality collapse [14]. In medical imaging, image-to-image transition frameworks are used for multiple tasks, including modality transition in which DDPMs and GANs showed promising performance [21,16,22,6,13]. Lyu et al [16] showed the superiority of diffusion models compared to GANs in this task for the conversion between MRI and Computed Tomography (CT) using a supervised framework (i.e., with pairs of data from both modalities). In unsupervised settings, Pan et al [22] developed a cycle-guided DDPM composed of two DDPMs that condition each other to generate synthetic images from two different MRI pulse sequences. Considering the achievements of these models in modality transition, which involves transitioning between distinct acquisition modalities, there is reason to anticipate their success in transitioning between other image types, such as analysis pipelines.

However, DDPMs are challenging to control when the objective is to generate images that maintain the intrinsic properties of the source images while transferring the extrinsic properties to the target domain. DDPMs are iterative generative models, i.e. they learn to model the transition from a Gaussian distribution to a target data distribution. Thus, data generated by the DDPM depend on the initial samples drawn from the Gaussian distribution, usually at random. Moreover, conditioning is often performed using a single target image or a one-hot encoded vector which might not represent the whole variability of the target domain.

In this work, we design a new unsupervised multi-domain DDPM framework in order to convert fMRI results (i.e. statistic maps) across pipelines, i.e. to convert a map from a pipeline (source domain) to the corresponding maps from another pipelines (target domains). Our contributions are as follow:

- We are the first to make the assumption that pipelines can be considered as extrinsic properties of statistic maps and can be transferred between maps.
- We extend the methods from [10,25] to build a classifier-conditional DDPM, called CCDDPM, in an unsupervised multi-domain transition framework to convert between multiple pipelines using a single model.
- We condition our CCDDPM using the latent space of a classifier trained to distinguish statistic maps between pipelines, a task previously unexplored.
- We propose a novel sampling strategy by selecting multiple target samples using a guided process based on clustering, to improve both transfer of target domain features while maintaining the source image properties.

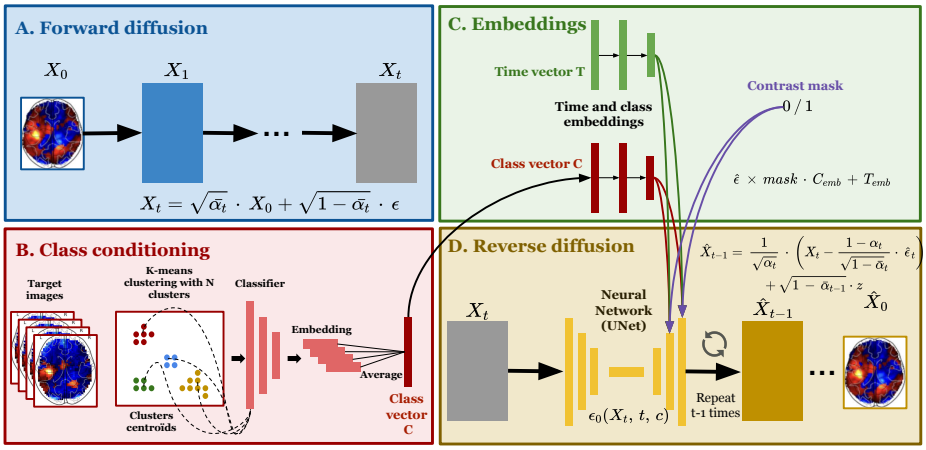

Fig. 1. Diagram of the workflow. During the forward diffusion (A), original maps X_0 are turned into X_t after t steps of noise addition ϵ. (B) Class conditioning uses latent vectors extracted from a classifier. These are averaged across N images, which are the centroids of N clusters identified using a K-Means algorithm. (C) Time and class are embedded using two Multi-Layers Perceptrons (MLP). A mask is applied to the class conditioning vector to jointly train an unconditional model with a pre-defined probability. (D) During the reverse diffusion, the neural network $\epsilon_\theta(X_t, t, c)$ learns to predict the noise added to the image and reconstructs X_{t-1} iteratively until $t = 0$.

– We compare CCDDPM to simpler DDPM models and also to a state-of-the-art multi-domain image-to-image transition model, starGAN [4], specifically implemented in 3D. We evaluate and compare the performance of each competitors on a broad set of metrics.

2 Classifier-Conditional Diffusion Model (CCDDPM)

Fig. 1 illustrates the model design. CCDDPM uses traditional diffusion processes [9] and the conditioning principles of [10]. We improve the sampling strategies by fixing the initial state of the DDPM with noisy source images to keep its intrinsic properties and by conditioning on the latent space of a classifier using multiple target images to represent the diversity of the target domain.

2.1 Diffusion processes (forward and reverse)

Diffusion models work by successively adding noise to the training data, and then learn to reverse the process to construct desired data samples from the noise. In the forward diffusion process (Fig. 1(A)), the source image X_0 is subjected to t steps of gradual noise ϵ addition to generate intermediate noisy versions of the image $\{X_0, X_1, ..., X_t\}$. The $t-th$ version of the image is expressed as:

$$X_t = \sqrt{\bar{\alpha}_t} * X_0 + \sqrt{1-\bar{\alpha}_t} \cdot \epsilon \quad \text{with} \quad \epsilon \sim N(0, I) \tag{1}$$

where α_t corresponds to fixed hyper-parameters between 0 and 1 related to the variance and $\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_t$, extracted from [9].

The reverse diffusion process (Fig. 1(D)) uses a neural network trained to predict the noise added to the image $\widehat{\epsilon_t} = \epsilon_\theta(X_t, t, C)$ at each time step t given X_t the noisy version of X_0, t the corresponding time step and C a conditioning vector. During training, the conditioning vector correspond to features of the true domain of the image, see 2.2). This way, the model learns to predict the noise added to the image while taking into account the domain of the image. This creates a multi-domain framework in which the model can predict the noise from any source image and to convert this image to each domain seen during training. Starting from X_t and using the predicted noise, the image X_{t-1} from previous step can be reconstructed using the following equation and we can reconstruct X_0 by repeating this process for t times:

$$\widehat{X}_{t-1} = \frac{1}{\sqrt{\alpha_t}} \cdot (X_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}} \cdot \widehat{\epsilon_t}) + \sqrt{1-\bar{\alpha}_{t-1}} \cdot z \quad \text{where} \quad z \sim N(0, I) \quad (2)$$

The equation is extracted from [9]. The network $\epsilon_\theta(X_t, t, C)$ is trained using a Mean Squared Error loss, $L_{MSE} := \mathbb{E}_{X,C,t,\epsilon\sim N(0,1)}[\| \epsilon_t - \widehat{\epsilon}_t \|_2^2]$.

2.2 Conditional diffusion

The general method of conditional diffusion follows the principles described in [10]. Timestep and conditioning are embedded using 2 MLP and infused with the neural network activations at a certain layer via $a_{L+1} = c_{emb} \cdot a_L + t_{emb}$ (Fig. 1(C)). An unconditional DDPM is trained along with the conditional one by setting a contrast mask m. This mask changes the conditioning vector to a null token $\emptyset$ with some probability $p_{\text{uncond.}}$, set as hyper-parameter. During sampling, we compute both conditional and unconditional noise prediction and we perform a linear combination of the two with a weight w to represent the strength of the conditional guidance using the equation from [10]:

$$\widehat{\epsilon_\theta}(X_t, t, C) = (1+w) \cdot \widehat{\epsilon_\theta}(X_t, t, C) - w \cdot \widehat{\epsilon_\theta}(X_t, t) \quad (3)$$

Classifier conditioning. In [10], the diffusion is conditioned using a one-hot encoding of the domain, which decreases the diversity of samples. In [25], a semantic encoder is used to guide sampling. In CCDDPM, we extend this idea by using a pre-trained Convolutional Neural Network (CNN) classifier that identifies the pipeline used in statistic maps to condition the model (Fig. 1(B)). The features are extracted just before the fully connected layer, to get a latent vector with the most important features that distinguish images across pipelines.

2.3 Sampling strategy for image-to-image transition

Multi-target images. To condition on the latent space of the classifier during sampling, target images must be selected. In [3], authors showed that conditioning on multiple images generates images that share coarse or fine features with

the target ones depending on the number of selected images. Selecting multiple target images to convert images between domains can help to generate images that represent the diversity of the target domain. In practice, the whole set of images available in the target domain could be used. This is impractical for large datasets and might lead the model to focus on specific patterns of the target domain if these are over-represented in the dataset. Here, we propose to use a K-Means algorithm [17] to identify N clusters of images in the target domain (Fig. 1(B)). Then, we extract the centroïd of these clusters and average their latent vector for conditioning. The number of cluster N must be chosen depending on the dataset and its diversity.

Source content preservation. To keep the intrinsic properties of the source image, [28] concatenated the source image along with random Gaussian noise to initialize the diffusion. Here, we propose to fix the initial state of the DDPM by directly using the forward diffusion process to generate a noisy version of the source image X_t (Equation 1). Then, the noisy source image is iteratively denoised using the predicted noise and the reverse diffusion process (Equation 2) with an additional conditioning on the target domain.

3 Experiments

3.1 Dataset

This work was performed using data from the Human Connectome Project [31]. Written informed consent was obtained from participants and the original study was approved by the Washington University Institutional Review Board. We agreed to the Open Access Data Use Terms. We used fMRI data from the motor task of the HCP Young Adult dataset [31] release S-1200 and computed the fMRI maps for the 1,080 participants using 4 analysis pipelines that differed in terms of software package (SPM [24] or FSL [12]) and presence or absence of the derivatives of the Hemodynamic Response Function (HRF) for the first-level analysis. We built 1,000 groups of 50 randomly-sampled participants, leading to 1,000 group-level statistic maps for each pipeline. In the following, we will refer to each pipeline as 'software-presence of hrf derivatives'. For instance, the pipeline using FSL and no HRF derivatives will be denoted by 'fsl-0'.

The selected group-level statistic maps were resampled to a size of 48 x 56 x 48 and masked using the intersection mask of all groups. The voxel values were normalized between -1 and 1 for each statistic maps using a min-max operation. The 1,000 groups were split into train, valid and test with a 90/8/2 ratio and all models were trained and evaluated on the same sets. Further investigation about possible data leakage across groups is provided in Supplementary Materials.

3.2 Implementation settings

Our code is available at : `https://anonymous.4open.science/r/style-transfer_diffusion-12B2`. Trained models are shared as Supplementary Materials.

The neural network used to predict the noise follows a simple U-Net architecture [27] with two downsampling and upsampling blocks with 3D convolutions layers and skip connections. The hyperparameters of the DDPMs are the following: $t = 500$ diffusion steps; linear noise schedule with variances in the range of $\beta_1 = 10^4$ and $\beta_t = 0.02$; batch size of 8 and learning rate of 1e-4. The weight w used to control the conditional guidance is optimized on the validation set by comparing $w = 0, w = 0.5$ and $w = 2$ and a value of 0.5 was found to give the best results in terms of Pearson's correlation coefficient between the target ground-truth and the generated image on this set. The model is implemented using PyTorch [23] and trained for 200 epochs on 1 GPU NVIDIA Tesla V100.

The CNN used to extract class conditional features is composed of five 3D convolution layers with 3D batch normalization and leaky rectified linear units (ReLU) activation functions, followed by a fully connected layer. The latent space corresponds to a $4,096$ flatten vector which is injected as conditioning to the U-Net. It is trained for 150 epochs using a learning rate of 1e-4 and a batch size of 64 on 1 GPU NVIDIA Tesla V100.

3.3 Evaluation metrics & competitors

We evaluate the performance using different metrics: Pearson's correlation (Corr.) in percent, Peak Signal to Noise Ratio (PSNR) and Inception Score (IS) [29], computed using the pipeline classifier described above. The first two metrics study the adequacy of generated images to the ground truth target, whereas IS explore the confidence of the conditional class predictions (quality) and the integral of the marginal probability of the predicted classes (diversity).

We compare CCDDPM to different variations of conditional DDPMs: conditioning using a one-hot encoding of the class, as in [10] and conditioning using the latent space of the classifier but only one target image selected randomly, inspired from [25]. We also implement a 3D version of the state-of-the-art starGAN [4] to compare the the performance of DDPMs with GANs.

In default settings, CCDDPM uses $N = 10$ target images. We also assess the impact of the number of target images for $N = 5, 10$ and 20. We also compare the selection process with a random sampling of target images and with a sampling based on the identification of images that are close to the source image using a K-Nearest Neighbors algorithm [19].

3.4 Results

In Table 1, we show the performance for 4 transfers, between two pipelines with: different HRF and different software (columns 1-4), same HRF and different software (columns 4-6) and, different HRF and same software (columns 6-8). Performance of CCDDPM are compared to the one in [10] and one inspired from [25] with classifier-conditioning and $N = 1$ target image selected randomly. Regarding the quality of the transfer to the target domain, all models succeed in changing the class identified by a pipeline classifier to the target domain. CCDDPM outperforms the other models (including starGAN) in terms of Inception

		fsl-1 → spm-0		spm-0 → fsl-1		fsl-1 → spm-1		fsl-1 → fsl-0	
	IS	Corr.	PSNR	Corr.	PSNR	Corr.	PSNR	Corr.	PSNR
Initial	*3.69*	*76.2*	*78.2*	*76.2*	*78.2*	*82.6*	*81.3*	*91.0*	*83.9*
One-hot [10]	3.66	83.8	77.2	75.0	79.4	78.7	77.7	81.1	79.5
N=1 [25]	3.70	85.5	79.0	77.8	80.0	79.9	78.0	82.8	80.2
StarGAN [4]	3.63	**90.5**	**81.9**	**87.5**	**84.2**	**87.6**	**81.8**	**91.5**	**85.0**
CCDDPM	**3.86**	86.1	79.4	79.0	80.7	81.2	78.9	84.1	80.6
N=5	3.86	86.4	79.8	78.7	80.6	81.2	79.4	84.5	80.9
N=20	3.87	86.1	79.5	79.2	80.7	81.3	79.2	83.9	80.9
N=5, random	3.89	86.5	79.4	79.1	80.4	82.0	79.2	84.2	80.2
N=10, random	3.86	86.5	79.2	79.0	80.2	81.8	79.4	84.3	80.8
N=20, random	3.85	86.7	79.1	79.3	80.6	81.5	79.4	84.4	80.7
N=10, KNN	3.75	84.9	79.3	78.7	80.0	81.6	79.1	83.6	80.7

Table 1. Performance associated with four transfers. IS means "Inception Score" across all transfers. Pearson's correlation (%) and Peak Signal to Noise Ration (PSNR) computed between generated and ground-truth target image for 20 images per transfer. *Initial* represents the metrics between the source image (before transfer) and the ground-truth target image. Default CCDDPM uses N=10. **Boldface marks the top model**.

Score. The use of a DDPM with classifier-conditioning and multiple target images seems to improve both quality and diversity of images, which might be explained by the non-determinism of sampling in DDPMs.

In terms of similarity to the ground-truth target image, CCDDPM outperforms the other DDPM models by up to 4% in correlations between target ground-truth and generated image compared to [10] for transfer spm-0 to fsl-1 and up to 3% for fsl-1 to spm-0. Even compared to the other best DDPM models in each task, CCDDPM is at least 0.6% better in terms of correlations. For the last two transfers in Table 1, our performance does not overpass the initial values between the source image and the ground-truth target image. These low performance could be explained by the difficulty of the models to learn differences between close pipelines. In Supplementary Figures, we show the performance of the pipeline classifier and compare the similarity of features.

However, the performance of starGAN [4] remain highly superior to the ones obtained with any DDPM, even for the last two transfers for which it overpasses the initial metrics. This superiority can be explained by the differences between frameworks: GAN-based methods use adversarial training and starGAN improves this by using a classifier loss and a cyclic-reconstruction loss. Moreover, GANs sampling rely on the source image directly and do not require to set an initial state, which might facilitate the source content preservation.

At the bottom of Table 1, we show the influence of the number of target images and of the selection methods. The number of images does not seem to impact the performance, correlations are very similar between $N = 5, N = 10$ and $N = 20$. Performing selection using K-Means algorithm does not seem to improve performance compared to a random selection, for any N values,

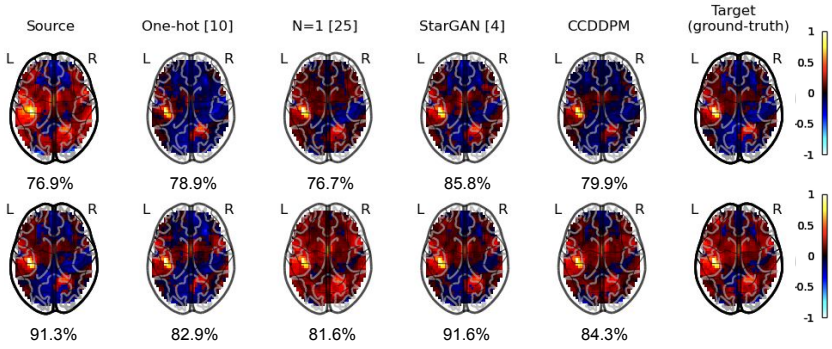

Fig. 2. Generated images for two transfer and different competitors: conditioning with one-hot encoding [10], with a classifier and N=1 [25], starGAN [4] and CCDPM. Correlation with target ground-truth are indicated below generated and source images.

probably due to the low diversity in our dataset. However, selection using a K-Nearest Neighbors (KNN) algorithm decreases the performance from 1.6%, meaning that the diversity of target images is beneficial for a good transfer.

The first row of Figure 2 illustrates a transfer between pipelines with different HRF and different software (spm-0 to fsl-1). The second row shows a transfer between pipelines with different HRF (fsl-1 to fsl-0). CCDDPM generates statistic maps close to the ground-truth for both transfer, representing the intrinsic properties of the map while modifying its extrinsic properties to the target domain. Maps generated using starGAN [4] remain closer to the target ground-truth, with more similar patterns, as stated by the similarity metrics.

4 Conclusion

We explore the ability to convert fMRI maps between pipelines using generative models. In particular, we provide a new DDPM unsupervised multi-domain image-to-image transition framework by conditioning on multiple target images in the latent space of a classifier. Our results show that images can be converted successfully, but with lower adequacy with the ground-truth target compared to GANs. In future work, we would like to focus on latent diffusion models to improve image quality and conservation of intrinsic properties of the image.

5 Acknowledgements

This work was funded by Region Bretagne (ARED MAPIS) and ANR project ANR-20-THIA-0018). Data were provided by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

References

1. Botvinik-Nezer, R., Holzmeister, F., Camerer, C.F., Dreber, A., Huber, J., Johannesson, M., Kirchler, M., Iwanir, R., Mumford, J.A., Adcock, R.A., Avesani, P., Baczkowski, B.M., Bajracharya, A., Bakst, L., Ball, S., Barilari, M.: Variability in the analysis of a single neuroimaging dataset by many teams. Nature **582**(7810), 84–88 (2020)
2. Carp, J.: On the plurality of (methodological) worlds: estimating the analytic flexibility of fMRI experiments. Front. Neurosci. p. 13 (2012)
3. Choi, J., Kim, S., Jeong, Y., Gwon, Y., Yoon, S.: ILVR: conditioning method for denoising diffusion probabilistic models. In: 2021 IEEE/CVF International Conference on Computer Vision, ICCV. pp. 14347–14356. IEEE (2021)
4. Choi, Y., Choi, M., Kim, M., Ha, J.W., Kim, S., Choo, J.: StarGAN: Unified Generative Adversarial Networks for Multi-domain Image-to-Image Translation. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 8789–8797. IEEE (2018)
5. Costafreda, S.: Pooling fMRI data: meta-analysis, mega-analysis and multi-center studies. Front. Neuroinform. **3** (2009)
6. Dai, X., Lei, Y., Fu, Y., Curran, W.J., Liu, T., Mao, H., Yang, X.: Multimodal MRI Synthesis Using Unified Generative Adversarial Networks. Medical physics **47**(12), 6343–6354 (2020)
7. Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: 27th International Conference on Neural Information Processing Systems - Volume 2. p. 2672–2680. NIPS'14 (2014)
8. Gorgolewski, K.J., Varoquaux, G., Rivera, G., Schwarz, Y., Ghosh, S.S., Maumet, C., Sochat, V.V., Nichols, T.E., Poldrack, R.A., Poline, J.B., Yarkoni, T., Margulies, D.S.: NeuroVault.org: a web-based repository for collecting and sharing unthresholded statistical maps of the human brain. Front. Neuroinform. (2015)
9. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. In: 34th International Conference on Neural Information Processing Systems. NIPS'20 (2020)
10. Ho, J., Salimans, T.: Classifier-Free Diffusion Guidance. In: "Deep Generative Models and Downstream Applications" Workshop@NeurIPS'21 (2021)
11. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5967–5976 (2017)
12. Jenkinson, M., Beckmann, C.F., Behrens, T.E.J., Woolrich, M.W., Smith, S.M.: FSL. NeuroImage **62**(2), 782–790 (2012)
13. Jin, C.B., Kim, H., Liu, M., Jung, W., Joo, S., Park, E., Ahn, Y.S., Han, I.H., Lee, J.I., Cui, X.: Deep CT to MR Synthesis Using Paired and Unpaired Data. Sensors **19**(10) (2019)
14. Li, H., Yang, Y., Chang, M., Chen, S., Feng, H., Xu, Z., Li, Q., Chen, Y.: Srdiff: Single image super-resolution with diffusion probabilistic models. Neurocomputing **479**, 47–59 (2022)
15. Liu, M., Breuel, T.M., Kautz, J.: Unsupervised image-to-image translation networks. In: Advances in Neural Information Processing Systems 30 (NIPS). pp. 700–708 (2017)
16. Lyu, Q., Wang, G.: Conversion Between CT and MRI Images Using Diffusion and Score-Matching Models. arXiv (2022)

17. MacQueen, J.: Some methods for classification and analysis of multivariate observations. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics, vol. 5.1, pp. 281–298 (1967)
18. Markiewicz, C.J., Gorgolewski, K.J., Feingold, F., Blair, R., Halchenko, Y.O., Miller, E., Hardcastle, N., Wexler, J., Esteban, O., Goncavles, M., Jwa, A., Poldrack, R.: The OpenNeuro resource for sharing of neuroscience data. eLife **10**, e71774 (2021)
19. Mucherino, A., Papajorgji, P.J., Pardalos, P.M.: k-nearest neighbor classification. In: Data Mining in Agriculture, pp. 83–106. Springer (2009)
20. Nichol, A.Q., Dhariwal, P.: Improved Denoising Diffusion Probabilistic Models. In: 38th International Conference on Machine Learning. pp. 8162–8171. PMLR (2021)
21. Özbey, M., Dalmaz, O., Dar, S.U.H., Bedel, H.A., Öztürk, S., Güngör, A., Çukur, T.: Unsupervised medical image translation with adversarial diffusion models. IEEE Trans. Medical Imaging **42**(12), 3524–3539 (2023)
22. Pan, S., Chang, C.W., Peng, J., Zhang, J., Qiu, R.L.J., Wang, T., Roper, J., Liu, T., Mao, H., Yang, X.: Cycle-guided Denoising Diffusion Probability Model for 3D Cross-modality MRI Synthesis. arXiv (2023)
23. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S.: Pytorch: An imperative style, high-performance deep learning library. In: Advances in Neural Information Processing Systems 32, pp. 8024–8035. Curran Associates, Inc. (2019)
24. Penny, W., Friston, K., Ashburner, J., Kiebel, S., Nichols, T.E.: Statistical Parametric Mapping: The Analysis of Functional Brain Images. Elsevier edn. (2011)
25. Preechakul, K., Chatthee, N., Wizadwongsa, S., Suwajanakorn, S.: Diffusion autoencoders: Toward a meaningful and decodable representation. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR. pp. 10609–10619. IEEE (2022)
26. Rolland, X., Maurel, P., Maumet, C.: Towards efficient fmri data re-use: can we run between-group analyses with datasets processed differently with spm ? In: ISBI 2022 - IEEE International Symposium on Biomedical Imaging. pp. 1–4 (2022)
27. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI). pp. 234–241 (2015)
28. Saharia, C., Chan, W., Chang, H., Lee, C.A., Ho, J., Salimans, T., Fleet, D.J., Norouzi, M.: Palette: Image-to-image diffusion models. In: SIGGRAPH '22: Special Interest Group on Computer Graphics and Interactive Techniques Conference. pp. 15:1–15:10. ACM (2022)
29. Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., Chen, X., Chen, X.: Improved Techniques for Training GANs. In: Lee, D., Sugiyama, M., Luxburg, U., Guyon, I., Garnett, R. (eds.) Advances in Neural Information Processing Systems. vol. 29. Curran Associates, Inc. (2016)
30. Sasaki, H., Willcocks, C.G., Breckon, T.P.: UNIT-DDPM: UNpaired Image Translation with Denoising Diffusion Probabilistic Models. arXiv (2021)
31. Van Essen, D.C., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T.E.J., WU-Minn HCP Consortium: The Human Connectome Project: a data acquisition perspective. NeuroImage **62**(4), 2222–2231 (2012)
32. Zhu, J., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: IEEE International Conference on Computer Vision, ICCV. pp. 2242–2251. IEEE Computer Society (2017)