CoLa-Diff: Conditional Latent Diffusion Model for Multi-modal MRI Synthesis

Lan Jiang[1], Ye Mao[2], Xiangfeng Wang[3], Xi Chen[4], and Chao Li[1,2,5(✉)]

[1] School of Science and Engineering, University of Dundee, Dundee, UK
[2] Department of Clinical Neurosciences, University of Cambridge, Cambridge, UK
[3] School of Computer Science and Technology, East China Normal University, Shanghai, China
[4] Department of Computer Science, University of Bath, Bath, UK
[5] School of Medicine, University of Dundee, Dundee, UK
cl647@cam.ac.uk

Abstract. MRI synthesis promises to mitigate the challenge of missing MRI modality in clinical practice. Diffusion model has emerged as an effective technique for image synthesis by modelling complex and variable data distributions. However, most diffusion-based MRI synthesis models are using a single modality. As they operate in the original image domain, they are memory-intensive and less feasible for multi-modal synthesis. Moreover, they often fail to preserve the anatomical structure in MRI. Further, balancing the multiple conditions from multi-modal MRI inputs is crucial for multi-modal synthesis. Here, we propose the first diffusion-based multi-modality MRI synthesis model, namely Conditioned Latent Diffusion Model (CoLa-Diff). To reduce memory consumption, we perform the diffusion process in the latent space. We propose a novel network architecture, e.g., similar cooperative filtering, to solve the possible compression and noise in latent space. To better maintain the anatomical structure, brain region masks are introduced as the priors of density distributions to guide diffusion process. We further present auto-weight adaptation to employ multi-modal information effectively. Our experiments demonstrate that CoLa-Diff outperforms other state-of-the-art MRI synthesis methods, promising to serve as an effective tool for multi-modal MRI synthesis.

Keywords: Multi-modal MRI · Medical image synthesis · Latent space · Diffusion models · Structural guidance

1 Introduction

Magnetic resonance imaging (MRI) is critical to the diagnosis, treatment, and follow-up of brain tumour patients [26]. Multiple MRI modalities offer complementary information for characterizing brain tumours and enhancing patient

L. Jiang and Y. Mao—Contribute equally in this work.

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14229, pp. 398–408, 2023.
https://doi.org/10.1007/978-3-031-43999-5_38

management [4,27]. However, acquiring multi-modality MRI is time-consuming, expensive and sometimes infeasible in specific modalities, e.g., due to the hazard of contrast agent [15]. Trans-modal MRI synthesis can establish the mapping from the known domain of available MRI modalities to the target domain of missing modalities, promising to generate missing MRI modalities effectively. The synthetic methods leveraging multi-modal MRI, i.e., many-to-one translation, have outperformed single-modality models generating a missing modality from another available modality, i.e., one-to-one translation [23,33]. Traditional multi-modal methods [21,22], e.g., sparse encoding-based, patch-based and atlas-based methods, rely on the alignment accuracy of source and target domains and are poorly scalable. Recent generative adversarial networks (GANs) and variants, e.g., MM-GAN [23], DiamondGAN [13] and ProvoGAN [30], have been successful based on multi-modal MRI, further improved by introducing multi-modal coding [31], enhanced architecture [7], and novel learning strategies [29].

Despite the success, GAN-based models are challenged by the limited capability of adversarial learning in modelling complex multi-modal data distributions [25] Recent studies have demonstrated that GANs' performance can be limited to processing and generating data with less variability [1]. In addition, GANs' hyperparameters and regularization terms typically require fine-tuning, which otherwise often results in gradient vanish and mode collapse [2].

Diffusion model (DM) has achieved state-of-the-art performance in synthesizing natural images, promising to improve MRI synthesis models. It shows superiority in model training [16], producing complex and diverse images [9,17], while reducing risk of modality collapse [12].For instance, Lyu et al. [14] used diffusion and score-marching models to quantify model uncertainty from Monte-Carlo sampling and average the output using different sampling methods for CT-to-MRI generation; Özbey et al. [19] leveraged adversarial training to increase the step size of the inverse diffusion process and further designed a cycle-consistent architecture for unpaired MRI translation.

However, current DM-based methods focus on one-to-one MRI translation, promising to be improved by many-to-one methods, which requires dedicated design to balance the multiple conditions introduced by multi-modal MRI. Moreover, as most DMs operate in original image domain, all Markov states are kept in memory [9], resulting in excessive burden. Although latent diffusion model (LDM) [20] is proposed to reduce memory consumption, it is less feasible for many-to-one MRI translation with multi-condition introduced. Further, diffusion denoising processes tend to change the original distribution structure of the target image due to noise randomness [14], rending DMs often ignore the consistency of anatomical structures embedded in medical images, leading to clinically less relevant results. Lastly, DMs are known for their slow speed of diffusion sampling [9,11,17], challenging its wide clinical application.

We propose a DM-based multi-modal MRI synthesis model, CoLa-Diff, which facilitates many-to-one MRI translation in latent space, and preserve anatomical structure with accelerated sampling. Our main contributions include:

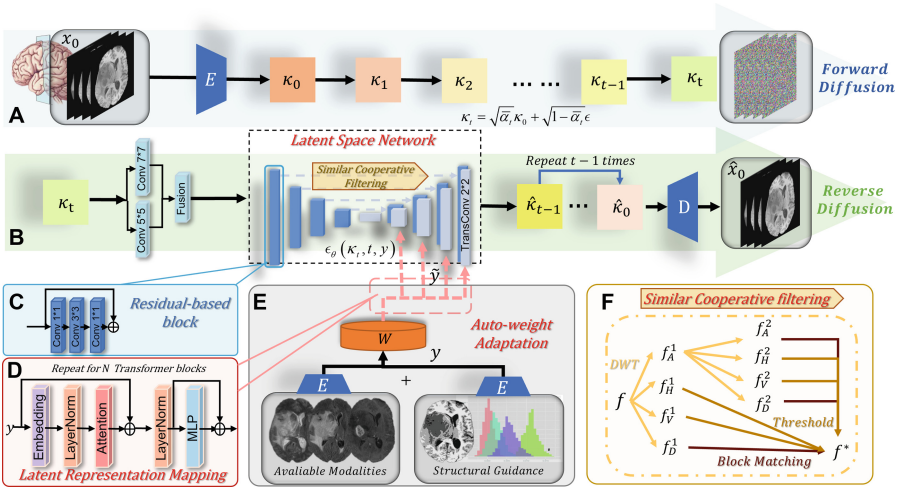

Fig. 1. Schematic diagram of CoLa-Diff. During the forward diffusion, Original images x_0 are compressed using encoder E to get κ_0, and after t steps of adding noise, the images turn into κ_t. During the reverse diffusion, the latent space network $\epsilon_\theta\left(\kappa_t, t, y\right)$ predicts the added noise, and other available modalities and anatomical masks as structural guidance are encoded to y, then processed by the auto-weight adaptation block W and embedded into the latent space network. Sampling from the distribution learned from the network gives $\hat{\kappa}_0$, then $\hat{\kappa}_0$ are decoded by D to obtain synthesized images.

- present a denoising diffusion probabilistic model based on multi-modal MRI. As far as we know, this is the first DM-based many-to-one MRI synthesis model.
- design a bespoke architecture, e.g., similar cooperative filtering, to better facilitate diffusion operations in the latent space, reducing the risks of excessive information compression and high-dimensional noise.
- introduce structural guidance of brain regions in each step of the diffusion process, preserving anatomical structure and enhancing synthesis quality.
- propose an auto-weight adaptation to balance multi-conditions and maximise the chance of leveraging relevant multi-modal information.

2 Multi-conditioned Latent Diffusion Model

Figure 1 illustrates the model design. As a latent diffusion model, CoLa-diff integrates multi-condition b from available MRI contrasts in a compact and low-dimensional latent space to guide the generation of missing modality $x \in \mathbb{R}^{H\times W\times 1}$. Precisely, b constitutes available contrasts and anatomical structure masks generated from the available contrasts. Similar to [9,20], CoLa-Diff invovles a forward and a reverse diffusion process. During forward diffusion, x_0 is encoded by E to produce κ_0, then subjected to T diffusion steps to gradually add

noise ϵ and generate a sequence of intermediate representations: $\{\kappa_0, \ldots, \kappa_T\}$. The t-th intermediate representation is denoted as κ_t, expressed as:

$$\kappa_t = \sqrt{\bar{\alpha}_t}\kappa_0 + \sqrt{1-\bar{\alpha}_t}\epsilon, \quad \text{with } \epsilon \sim \mathcal{N}(0, \mathbf{I}) \tag{1}$$

where $\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i$, α_i denotes hyper-parameters related to variance.

The reverse diffusion is modelled by a latent space network with parameters θ, inputting intermediate perturbed feature maps κ_t and y (compressed b) to predict noise level $\epsilon_\theta(\kappa_t, t, y)$ for recovering feature maps $\hat{\kappa}_{t-1}$ from previous,

$$\hat{\kappa}_{t-1} = \sqrt{\bar{\alpha}_{t-1}}\left(\frac{\kappa_t - \sqrt{1-\bar{\alpha}_t}\cdot\epsilon_\theta(\kappa_t, t, y)}{\sqrt{\bar{\alpha}_t}}\right) + \sqrt{1-\bar{\alpha}_{t-1}}\cdot\epsilon_\theta(\kappa_t, t, y) \tag{2}$$

To enable effective learning of the underlying distribution of κ_0, the noise level needs to be accurately estimated. To achieve this, the network employs similar cooperative filtering and auto-weight adaptation strategies. $\hat{\kappa}_0$ is recovered by repeating Eq. 2 process for t times, and decoding the final feature map to generate synthesis images $\hat{x_0}$.

2.1 Latent Space Network

We map multi-condition to the latent space network for guiding noise prediction at each step t. The mapping is implemented by N transformer blocks (Fig. 1 (D)), including global self-attentive layers, layer-normalization and position-wise MLP. Following the latent diffusion model (LDM) [20], the network $\epsilon_\theta(\kappa_t, t, y)$ is trained to predict the noise added at each step using

$$\mathcal{L}_{\mathrm{E}} := \mathbb{E}_{E(x),y,\epsilon\sim\mathcal{N}(0,1),t}\left[\|\epsilon - \epsilon_\theta(\kappa_t, t, y)\|_2^2\right] \tag{3}$$

To mitigate the excessive information losses that latent spaces are prone to, we replace the simple convolution operation with a residual-based block (three sequential convolutions with kernels $1*1$, $3*3$, $1*1$ and residual joins [8]), and enlarge the receptive field by fusion ($5*5$ and $7*7$ convolutions followed by AFF [6]) in the down-sampling section. Moreover, to reduce high-dimensional noise generated in the latent space, which can significantly corrupt the quality of multi-modal generation. We design a similar cooperative filtering detailed below.

Similar Cooperative Filtering. The approach has been devised to filter the downsampled features, with each filtered feature connected to its respective upsampling component (shown in Fig. 1 (F)). Given f, which is the downsampled feature of κ_t, suppose the 2D discrete wavelet transform ϕ [24] decomposes the features into low frequency component $f_A^{(i)}$ and high frequency components $f_H^{(i)}$, $f_V^{(i)}$, $f_D^{(i)}$, keep decompose $f_A^{(i)}$, where i is the number of wavelet transform layers. Previous work [5] has shown to effectively utilize global information by considering similar patches. However, due to its excessive compression, it is less suitable for LDM. Here, we group the components and further filter by similar

block matching δ [18] or thresholding γ, use the inverse wavelet transform $\phi^{-1}(\cdot)$ to reconstruct the denoising results, given f^*.

$$f^* = \phi^{-1}(\delta(f_A^{(i)}), \delta(\sum_{j=1}^{i} f_D^{(i)}), \gamma(\sum_{j=1}^{i} f_H^{(i)}), \gamma(\sum_{j=1}^{i} f_V^{(i)})) \tag{4}$$

2.2 Structural Guidance

Unlike natural images, medical images encompass rich anatomical information. Therefore, preserving anatomical structure is crucial for MRI generation. However, DMs often corrupt anatomical structure, and this limitation could be due to the learning and sampling processes of DMs that highly rely on the probability density function [9], while brain structures by nature are overlapping in MRI density distribution and even more complicated by pathological changes.

Previous studies show that introducing geometric priors can significantly improve the robustness of medical image generation. [3,28]. Therefore, we hypothesize that incorporating structural prior could enhance the generation quality with preserved anatomy. Specifically, we exploit FSL-FAST [32] tool to segment four types of brain tissue: white matter, grey matter, cerebrospinal fluid, and tumour. The generated tissue masks and inherent density distributions (Fig. 1 (E)) are then used as a condition y_i to guide the reverse diffusion.

The combined loss function for our multi-conditioned latent diffusion is defined as

$$\mathcal{L}_{\mathrm{MCL}} := \mathcal{L}_{\mathrm{E}} + \mathcal{L}_{\mathrm{KL}} \tag{5}$$

where KL is the KL divergence loss to measure similarity between real q and predicted p_θ distributions of encoded images.

$$\mathcal{L}_{\mathrm{KL}} := \sum_{j=1}^{T-1} D_{KL}\left(q\left(\kappa_{j-1} \mid \kappa_j, \kappa_0\right) \| p_\theta\left(\kappa_{j-1} \mid \kappa_j\right)\right) \tag{6}$$

where D_{KL} is the KL divergence function.

2.3 Auto-Weight Adaptation

It is critical to balance multiple conditions, maximizing relevant information and minimising redundant information. For encoded conditions $y \in \mathbb{R}^{h \times w \times c}$, c is the number of condition channels. Set the value after auto-weight adaptation to $\tilde{y}$, the operation of this module is expressed as (shown in Fig. 1 (E))

$$\tilde{y} = F(y|\mu, \nu, o), \quad \text{with } \mu, \nu, o \in \mathbb{R}^c \tag{7}$$

The embedding outputs are adjusted by embedding weight μ. The auto-activation is governed by the learnable weight ν and bias o. y_c indicates each channel of y, where $y_c = [y_c^{m,n}]_{h \times w} \in R^{h \times w}$, $y_c^{m,n}$ is the eigenvalue at position (m, n) in channel c. We use large receptive fields and contextual embedding

to avoid local ambiguities, providing embedding weight $\mu = [\mu_1, \mu_2..., \mu_c]$. The operation G_c is defined as:

$$G_c = \mu_c \left\| y_c \right\|_2 = \mu_c \left\{ \left[\sum_{m=1}^{h} \sum_{n=1}^{w} (y_c^{m,n})^2 \right] + \varpi \right\}^{\frac{1}{2}} \tag{8}$$

where ϖ is a small constant added to the equation to avoid the issue of derivation at the zero point. The normalization method can establish stable competition between channels, $\mathbf{G} = \{G_c\}_{c=1}^{S}$. We use L_2 normalization for cross-channel operations:

$$\hat{G}_c = \frac{\sqrt{S} G_c}{\|\mathbf{G}\|_2} = \frac{\sqrt{S} G_c}{\left[\left(\sum_{c=1}^{S} G_c^2 \right) + \varpi \right]^{\frac{1}{2}}} \tag{9}$$

where S denotes the scale. We use an activation mechanism for updating each channel to facilitate the maximum utilization of each condition during diffusion model training, and further enhance the synthesis performance. Given the learnable weight $\nu = [\nu_1, \nu_2, ..., \nu_c]$ and bias $\mathbf{o} = [o_1, o_2, ..., o_c]$ we compute

$$\tilde{y}_c = y_c[1 + S(\nu_c \hat{G}_c + o_c)] \tag{10}$$

which gives new representations $\tilde{y}_c$ of each compressed conditions after the automatic weighting. $S(\cdot)$ denotes the Sigmoid activation function.

3 Experiments and Results

3.1 Comparisons with State-of-the-Art Methods

Datasets and Baselines. We evaluated CoLa-Diff on two multi-contrast brain MRI datasets: BRATS 2018 and IXI datasets. The BRATS 2018 contains MRI scans from 285 glioma patients. Each includes four modalities: T1, T2, T1ce, and FLAIR. We split them into (190:40:55) for training/validation/testing. For each subject, we automatically selected axial cross-sections based on the perceptible effective area of the slices, and then cropped the selected slices to a size of 224×224. The IXI[1] dataset consists of 200 multi-contrast MRIs from healthy brains, plit them into (140:25:35) for training/validation/testing. For preprocessing, we registered T2- and PD-weighted images to T1-weighted images using FSL-FLIRT [10], and other preprocessing are identical to the BRATS 2018.

We compared CoLa-Diff with four state-of-the-art multi-modal MRI synthesis methods: MM-GAN [23], Hi-Net [33], ProvoGan [30] and LDM [20].

Implementation Details. Our code is publicly available at https://github.com/SeeMeInCrown/CoLa_Diff_MultiModal_MRI_Synthesis. The hyperparameters of CoLa-Diff are defined as follows: diffusion steps to 1000; noise schedule to linear; attention resolutions to $32, 16, 8$; batch size to 8, learning rate to $9.6e-5$.

[1] https://brain-development.org/ixi-dataset/.

The noise variances were in the range of $\beta_1 = 10^{-4}$ and $\beta_T = 0.02$. An exponential moving average (EMA) over model parameters with a rate of 0.9999 was employed. The model is trained on 2 NVIDIA RTX A5000, 24 GB with Adam optimizer on PyTorch. An acceleration method [11] based on knowledge distillation was applied for fast sampling.

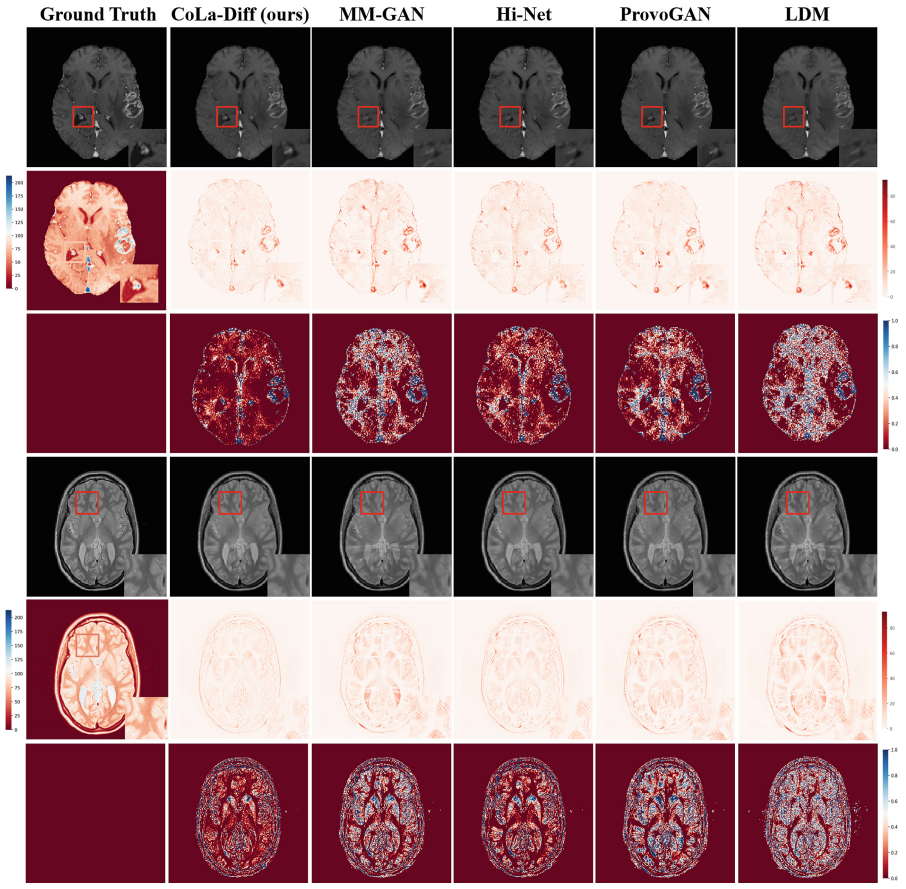

Fig. 2. Visualization of synthesized images, detail enlargements (row 1 and 4), corresponding error maps (row 2 and 5) and uncertainty maps (row 3 and 6).

Quantitative Results. We performed synthesis experiments for all modalities, with each modality selected as the target modality while remaining modalities and the generated region masks as conditions. Seven cases were tested in two datasets (Table 1). The results show that CoLa-Diff outperforms other models by up to 6.01 dB on PSNR and 5.74% on SSIM. Even when compared to the best of other models in each task, CoLa-Diff is a maximum of 0.81 dB higher in PSNR and 0.82% higher in SSIM.

Table 1. Performance in BRATS (top) and IXI (bottom). PSNR (dB) and SSIM (%) are listed as mean±std in the test set. **Boldface** marks the top models.

Model (BRATS 2018)	T2+T1ce+FLAIR →T1		T1+T1ce+FLAIR →T2		T2+T1+FLAIR →T1ce		T2+T1ce+T1 →FLAIR	
	PSNR	SSIM%	PSNR	SSIM%	PSNR	SSIM%	PSNR	SSIM%
MM-GAN	25.78±2.16	90.67±1.45	26.11±1.62	90.58±1.39	26.30±1.91	91.22±2.08	24.09±2.14	88.32±1.98
Hi-Net	27.42±2.58	93.46±1.75	25.64±2.01	92.59±1.42	27.02±1.26	93.35±1.34	25.87±2.82	91.22±2.13
ProvoGAN	27.79±4.42	93.51±3.16	26.72±2.87	92.98±3.91	29.26±2.50	93.96±2.34	25.64±2.77	90.42±3.13
LDM	24.55±2.62	88.34±2.51	24.79±2.67	88.47±2.60	25.61±2.48	89.18±2.55	23.12±3.16	86.90±3.24
CoLa-Diff (Ours)	**28.26±3.13**	**93.65±3.02**	**28.33±2.27**	**93.80±2.75**	**29.35±2.40**	**94.18±2.46**	**26.68±2.74**	**91.89±3.11**

Model (IXI)	T1+T2 →PD		T2+PD →T1		T1+PD →T2	
	PSNR	SSIM%	PSNR	SSIM%	PSNR	SSIM%
MM-GAN	30.61±1.64	95.42±1.90	27.32±1.70	92.35±1.58	30.87±1.75	94.68±1.42
Hi-Net	31.79±2.26	96.51±2.03	28.89±1.43	93.78±1.31	32.58±1.85	96.54±1.74
ProvoGAN	29.93±3.11	94.62±2.40	24.21±2.63	90.46±3.58	29.19±3.04	94.08±3.87
LDM	27.36±2.48	91.52±2.39	24.19±2.51	88.75±2.47	27.04±2.31	91.23±2.24
CoLa-Diff (Ours)	**32.24±2.95**	**96.95±2.26**	**30.20±2.38**	**94.49±2.15**	**32.86±2.83**	**96.57±2.27**

Qualitative Results. The first three and last three rows in Fig. 2 illustrate the synthesis results of T1ce from BRATS and PD from the IXI, respectively. From the generated images, we observe that CoLa-Diff is most comparable to the ground truth, with fewer errors shown in the heat maps. The synthesis uncertainty for each region is derived by performing 100 generations of the same slice and calculating the pixel-wise variance. From the uncertainty maps, CoLa-Diff is more confident in synthesizing the gray and white matter over other comparison models. Particularly, CoLa-Diff performs better in generating complex brain sulcus and tumour boundaries. Further, CoLa-Diff could better maintain the anatomical structure over comparison models.

3.2 Ablation Study and Multi-modal Exploitation Capabilities

We verified the effectiveness of each component in CoLa-Diff by removing them individually. We experimented on BRATS T1+T1ce+FLAIR→T2 task with four absence scenarios (Table 2 top). Our results show that each component contributes to the performance improvement, with Auto-weight adaptation bringing a PSNR increase of 1.9450dB and SSIM of 4.0808%.

To test the generalizability of CoLa-Diff under the condition of varied inputs, we performed the task of generating T2 on two datasets with progressively increasing input modalities (Table 2 bottom). Our results show that our model performance increases with more input modalities: SSIM has a maximum uplift value of 1.9603, PSNR rises from 26.6355 dB to 28.3126 dB in BRATS; from 32.164 dB to 32.8721 dB in IXI. The results could further illustrate the ability of CoLa-Diff to exploit multi-modal information.

Table 2. Ablation of four individual components (First four lines) and Multi-modal information utilisation (Last three lines). **Boldface** marks the best performing scenarios on each dataset.

	PSNR	SSIM%
w/o Modified latent diffusion network	27.1074	90.1268
w/o Structural guidance	27.7542	91.4865
w/o Auto-weight adaptation	26.3896	89.7129
w/o Similar cooperative filtering	27.9753	92.1584
T1 (BRATS)	26.6355	91.7438
T1+T1ce (BRATS)	27.3089	92.9772
T1+T1ce+Flair (BRATS)	**28.3126**	**93.7041**
T1 (IXI)	32.1640	96.0253
T1+PD (IXI)	**32.8721**	**96.5932**

4 Conclusion

This paper presents CoLa-Diff, a DM-based multi-modal MRI synthesis model with a bespoke design of network backbone, similar cooperative filtering, structural guidance and auto-weight adaptation. Our experiments support that CoLa-Diff achieves state-of-the-art performance in multi-modal MRI synthesis tasks. Therefore, CoLa-Diff could serve as a useful tool for generating MRI to reduce the burden of MRI scanning and benefit patients and healthcare providers.

References

1. Bau, D., et al.: Seeing what a GAN cannot generate. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 4502–4511 (2019)
2. Berard, H., Gidel, G., Almahairi, A., Vincent, P., Lacoste-Julien, S.: A closer look at the optimization landscapes of generative adversarial networks. arXiv preprint: arXiv:1906.04848 (2019)
3. Brooksby, B.A., Dehghani, H., Pogue, B.W., Paulsen, K.D.: Near-infrared (NIR) tomography breast image reconstruction with a priori structural information from MRI: algorithm development for reconstructing heterogeneities. IEEE J. Sel. Top. Quantum Electron. **9**(2), 199–209 (2003)
4. Cherubini, A., Caligiuri, M.E., Péran, P., Sabatini, U., Cosentino, C., Amato, F.: Importance of multimodal MRI in characterizing brain tissue and its potential application for individual age prediction. IEEE J. Biomed. Health Inform. **20**(5), 1232–1239 (2016)
5. Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Image restoration by sparse 3D transform-domain collaborative filtering. In: Image Processing: Algorithms and Systems VI, vol. 6812, pp. 62–73. SPIE (2008)
6. Dai, Y., Gieseke, F., Oehmcke, S., Wu, Y., Barnard, K.: Attentional feature fusion. CoRR abs/2009.14082 (2020)
7. Dalmaz, O., Yurt, M., Çukur, T.: ResViT: residual vision transformers for multimodal medical image synthesis. IEEE Trans. Med. Imaging **41**(10), 2598–2614 (2022)
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016)

9. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. In: Advances in Neural Information Processing Systems, vol. 33, pp. 6840–6851 (2020)
10. Jenkinson, M., Smith, S.: A global optimisation method for robust affine registration of brain images. Med. Image Anal. **5**(2), 143–156 (2001)
11. Kong, Z., Ping, W.: On fast sampling of diffusion probabilistic models. arXiv preprint: arXiv:2106.00132 (2021)
12. Li, H., et al.: SRDiff: single image super-resolution with diffusion probabilistic models. Neurocomputing **479**, 47–59 (2022)
13. Li, H., et al.: DiamondGAN: unified multi-modal generative adversarial networks for MRI sequences synthesis. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11767, pp. 795–803. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32251-9_87
14. Lyu, Q., Wang, G.: Conversion between CT and MRI images using diffusion and score-matching models. arXiv preprint: arXiv:2209.12104 (2022)
15. Merbach, A.S., Helm, L., Toth, E.: The Chemistry of Contrast Agents in Medical Magnetic Resonance Imaging. John Wiley & Sons, Hoboken (2013)
16. Müller-Franzes, G., et al.: Diffusion probabilistic models beat gans on medical images. arXiv preprint: arXiv:2212.07501 (2022)
17. Nichol, A.Q., Dhariwal, P.: Improved denoising diffusion probabilistic models. In: International Conference on Machine Learning, pp. 8162–8171. PMLR (2021)
18. Ourselin, S., Roche, A., Prima, S., Ayache, N.: Block matching: a general framework to improve robustness of rigid registration of medical images. In: Delp, S.L., DiGoia, A.M., Jaramaz, B. (eds.) MICCAI 2000. LNCS, vol. 1935, pp. 557–566. Springer, Heidelberg (2000). https://doi.org/10.1007/978-3-540-40899-4_57
19. Özbey, M., et al.: Unsupervised medical image translation with adversarial diffusion models. arXiv preprint: arXiv:2207.08208 (2022)
20. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10684–10695 (2022)
21. Roy, S., Carass, A., Prince, J.: A compressed sensing approach for MR tissue contrast synthesis. In: Székely, G., Hahn, H.K. (eds.) IPMI 2011. LNCS, vol. 6801, pp. 371–383. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-22092-0_31
22. Roy, S., Carass, A., Prince, J.L.: Magnetic resonance image example-based contrast synthesis. IEEE Trans. Med. Imaging **32**(12), 2348–2363 (2013)
23. Sharma, A., Hamarneh, G.: Missing MRI pulse sequence synthesis using multi-modal generative adversarial network. IEEE Trans. Med. Imaging **39**(4), 1170–1183 (2019)
24. Shensa, M.J., et al.: The discrete wavelet transform: wedding the a trous and Mallat algorithms. IEEE Trans. Signal Process. **40**(10), 2464–2482 (1992)
25. Thanh-Tung, H., Tran, T.: Catastrophic forgetting and mode collapse in GANs. In: 2020 International Joint Conference on Neural Networks (IJCNN), pp. 1–10. IEEE (2020)
26. Vlaardingerbroek, M.T., Boer, J.A.: Magnetic Resonance Imaging: Theory and Practice. Springer Science & Business Media, Cham (2013)
27. Wei, Y., et al.: Multi-modal learning for predicting the genotype of glioma. IEEE Trans. Med. Imaging (2023)
28. Yu, B., Zhou, L., Wang, L., Shi, Y., Fripp, J., Bourgeat, P.: Ea-GANs: edge-aware generative adversarial networks for cross-modality MR image synthesis. IEEE Trans. Med. Imaging **38**(7), 1750–1762 (2019)

29. Yu, Z., Han, X., Zhang, S., Feng, J., Peng, T., Zhang, X.Y.: MouseGAN++: unsupervised disentanglement and contrastive representation for multiple MRI modalities synthesis and structural segmentation of mouse brain. IEEE Trans. Med. Imaging **42**, 1197–1209 (2022)
30. Yurt, M., Özbey, M., Dar, S.U., Tinaz, B., Oguz, K.K., Çukur, T.: Progressively volumetrized deep generative models for data-efficient contextual learning of MR image recovery. Med. Image Anal. **78**, 102429 (2022)
31. Zhan, B., Li, D., Wu, X., Zhou, J., Wang, Y.: Multi-modal MRI image synthesis via GAN with multi-scale gate mergence. IEEE J. Biomed. Health Inform. **26**(1), 17–26 (2022)
32. Zhang, Y., Brady, M., Smith, S.: Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE Trans. Med. Imaging **20**(1), 45–57 (2001)
33. Zhou, T., Fu, H., Chen, G., Shen, J., Shao, L.: Hi-Net: hybrid-fusion network for multi-modal MR image synthesis. IEEE Trans. Med. Imaging **39**(9), 2772–2781 (2020)