Contrast Representation Learning from Imaging Parameters for Magnetic Resonance Image Synthesis

Honglin Xiong[1]†, Yu Fang[1]†, Kaicong Sun[1], Yulin Wang[2], Xiaopeng Zong[1], Weijun Zhang[3], Qian Wang[1,4](✉)

[1] School of Biomedical Engineering & State Key Laboratory of Advanced Medical Materials and Devices, ShanghaiTech University, Shanghai, China
qianwang@shanghaitech.edu.cn
[2] School of Biomedical Engineering, Hainan University, Haikou, China
[3] Shanghai United Imaging Healthcare Co., Ltd., Shanghai, China
[4] Shanghai Clinical Research and Trial Center, Shanghai, China

Abstract. Magnetic Resonance Imaging (MRI) is a widely used non-invasive medical imaging technique that provides excellent contrast for soft tissues, making it invaluable for diagnosis and intervention. Acquiring multiple contrast images is often desirable for comprehensive evaluation and precise disease diagnosis. However, due to technical limitations, patient-related issues, and medical conditions, obtaining all desired MRI contrasts is not always feasible. Cross-contrast MRI synthesis can potentially address this challenge by generating target contrasts based on existing source contrasts. In this work, we propose Contrast Representation Learning (CRL), which explores the changes in MRI contrast by modifying MR sequences. Unlike generative models that treat image generation as an end-to-end cross-domain mapping, CRL aims to uncover the complex relationships between contrasts by embracing the interplay of imaging parameters within this space. By doing so, CRL enhances the fidelity and realism of synthesized MR images, providing a more accurate representation of intricate details. Experimental results on the Fast Spin Echo (FSE) sequence demonstrate the promising performance and generalization capability of CRL, even with limited training data. Moreover, CRL introduces a perspective of considering imaging parameters as implicit coordinates, shedding light on the underlying structure governing contrast variation in MR images. Our code is available at `https://github.com/xionghonglin/CRL_MICCAI_2024`.

Keywords: Image translation · Cross-Contrast synthesis · MRI sequences.

1 Introduction

Magnetic Resonance Imaging (MRI) plays a crucial role in contemporary medical diagnosis and intervention due to its non-invasive nature and excellent soft

† These authors contributed equally to this work.

tissue contrast. By adjusting imaging parameters during the scanning process, MR images depicting the same anatomical structure but with varying contrasts can be acquired. In clinical scenarios, obtaining multiple contrast MR images is often necessary for comprehensive evaluation of pathological conditions and precise diagnoses. The fusion of images with different tissue contrasts facilitates accurate treatment planning. However, challenges may arise during the MRI scan that restrict the successful acquisition of all desired contrast images. Factors such as the need to skip or shorten sequences for timely completion, technical limitations (hardware or software issues), and patient-related issues, such as motion artifacts, claustrophobia, or specific medical conditions, can impede the seamless acquisition of MRI sequences. In some instances, medical conditions or physical limitations may further restrict the acquisition of certain MRI contrasts.

To address these challenges, the synthesis of various contrast MRI has become a significant area of interest for researchers. MR image synthesis involves generating a "target" contrast image based on an existing "source" contrast image, which is a cross-domain image translation problem. This process entails learning the underlying relationship between the source and target contrasts, enabling the generation of realistic target contrasts that correspond to the sources. Synthesizing target contrasts aids clinicians and researchers in obtaining a more comprehensive understanding of anatomical structures or pathologies, particularly for tasks requiring all contrasts as inputs in automatic multi-modal image analysis. The field of MR image synthesis has witnessed substantial progress, incorporating traditional methods like atlas-based approaches [11] and random forest techniques [1,6]. More recently, deep learning methods [2,7,8,12,16,17] have revolutionized MR image synthesis, leveraging powerful neural network architectures and extensive training datasets to achieve remarkable results. The prevailing approach in existing studies involves treating the generation of MR images as an end-to-end cross-domain mapping. While this method is widely considered universal and reasonable, it typically involves mutual mappings between two to four MRI contrasts [3,9,10,14,15]. Despite their prevalence, we posit that directly using deep learning models to map between multiple discrete domains may not accurately capture the intricate variations in contrast observed in MR images. More importantly, these models are limited to translations between fixed contrasts and cannot synthesize MR images with unseen contrasts that were not present in the training data.

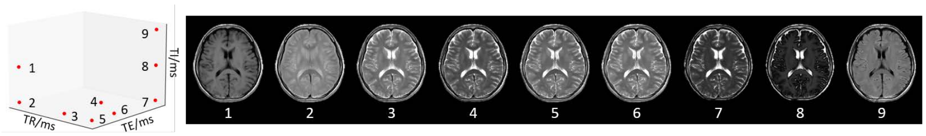

Fig. 1. Demonstration of all the actually acquired MR sequences in our paper. Points in the left show the sequence locations in the parameter space. Detailed imaging parameters are provided in Table 1.

In this research, we introduce Contrast Representation Learning (CRL), an approach that transcends the conventional paradigm of direct cross-domain mapping. Our hypothesis is rooted in the idea that MR images with different contrasts exist within a high-dimensional manifold. Within this manifold, subtle nuances in contrast are not adequately represented by a direct mapping between discrete domains. Instead, we propose that the intricate relationships between contrasts are better understood through the exploration of a high-dimensional space. In this latent space, the coordinates are implicitly expressed by the imaging parameters of the scanning sequence, revealing a richer understanding of the underlying structure governing contrast variation in MR images. To capture the implicit relationship between imaging parameters and image contrast within the network, we conducted a series of experiments on the Fast Spin Echo (FSE) sequence by varying parameters such as repetition time (TR), echo time (TE), and inversion time (TI), resulting in a collection of diverse sequences. Through the joint learning of multiple sequences, our model fits the high-dimensional manifold formed by multiple sequence contrasts, as well as the coordinates of each contrast expressed by imaging parameters.

CRL demonstrates its prowess by delivering impressive multi-contrast generation results across datasets, with training on just ten subjects. This underscores the robust and adaptable nature of the CRL framework, enabling effective contrast generation even with a small subset of subjects. Notably, CRL can generate contrast images formed by new imaging parameters that are not encountered during training. This has the potential to provide clinical practitioners with optimal contrast tailored to individual subjects.

2 Method

In this section, we first formulate the problem definition of our imaging parameter based MR image synthesis. Then we describe our network and the proposed contrast representation learning in details.

2.1 Problem Definition

Generally, the MRI signal intensity S from a SE (Spin Echo) sequence can be approximated as:

$$S_{SE} = f(k, t, p), \tag{1}$$

where k represents parameters including scaling factor and spin density. t denotes the T1 and T2 values of the scanning object, and p are the imaging parameters (e.g., TR, TE, TI). Since our goal is to synthesize the target sequence from the source sequence, and p controls the image contrast, we define a decoding function f_θ and an encoding function E_ϕ with θ and ϕ being the parameters, respectively. The relation between two contrast images S_x and S_y can be modeled as:

$$S_y = f_\theta(E_\phi(S_x), p_x, p_y), \tag{2}$$

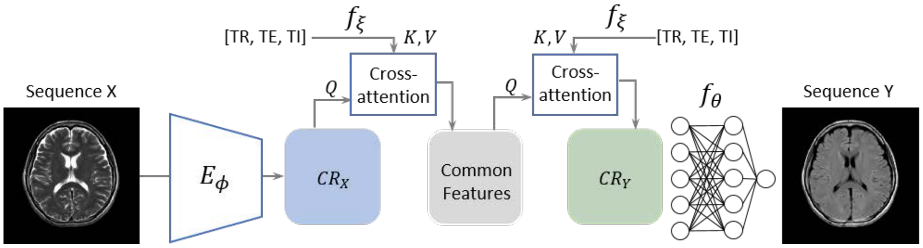

Fig. 2. Overview of our method. The encoder (E_ϕ) extracts the contrast representation CR_x from the source modality X. The source sequence parameters are mapped to a high-dimensional vector and cross-attended with CR_x to obtain common features across sequences. Using these common features and the target sequence parameter mapping, the target contrast representation CR_y is obtained, from which the decoder f_θ synthesizes the target modality Y.

where p_x and p_y denote the imaging parameters of the source and target sequence, respectively. The encoding function takes the source image S_x and maps it to its contrast representation in a high-dimensional space. To ease the interaction between contrast representation and imaging parameters, we map the three-dimensional sequence parameters (TR, TE, TI) into a high-dimensional space using a mapping function f_ξ:

$$p_\xi = f_\xi(TR, TE, TI) \tag{3}$$

Synthesizing the target sequence from a source sequence aims to learn the parameters of the encoder E_ϕ, the decoder f_θ, and the mapping function f_ξ.

2.2 Contrast Representation Learning

An overview of our method is illustrated in Fig. 2. Our model consists of an encoder E_ϕ, a decoder function f_θ, and an embedding function f_ξ. The encoder E_ϕ maps the source sequence X to its contrast representation CR_X. To generate all the contrasts controlled by imaging parameters, we construct the common features of all contrasts by performing cross-attention between CR_X and the high-dimensional embedding of p_x. The common features are then modulated by the embedding of the target imaging parameter p_y to construct the target contrast representation CR_Y, which is subsequently mapped to the target sequence Y by the decoder function f_θ. The learning of the relationship between contrast-related features and unrelated features is termed contrast representation learning.

Specifically, all sequences used for training are individually input into the same encoder and encoded as contrast representations. We use a ResNet consisting of 32-layer res-blocks as E_ϕ. The 3-dimensional imaging parameters are mapped to a 256-dimensional vector by a 4-layer MLP as f_ξ. Subsequently, all features serve as queries, and the corresponding imaging parameter embeddings act as keys and values, participating in cross-attention and being mapped onto the common features. We choose 8 heads for the multi-head attention in the cross-attention. Following that, the common features randomly select imaging parameters from another sequence as the condition and map it to the corresponding contrast representation. The decoding function is then applied to obtain the respective image. For the decoding function, we employ a 5-layer MLP with dimensions [256, 256, 256, 256, 1].

To establish the common features, we enforce structural consistency constraints on the features of all sequences within the latent space. Specifically, for the i-th and i'-th (input) sequences of a given subject, the common features from them are $\mathcal{Z}_i$ and $\mathcal{Z}_{i'}$, respectively. The classical Structural Similarity Index Measure (SSIM) is often used to gauge the similarity of two images [13]. We only utilize the structural component in SSIM to enforce the structural consistency between the two feature maps of $\mathcal{Z}_i$ and $\mathcal{Z}_{i'}$, while the luminance and contrast components in SSIM are not used. The structural consistency loss function is formulated as:

$$\mathcal{L}_s = -\sum_{i \neq i'} SSIM_s(\mathcal{Z}_i, \mathcal{Z}_{i'}), \tag{4}$$

where s indicates that only the structural component in SSIM calculation is preserved.

In addition to the loss for building the common features, we need another loss for the synthesized image. Let $\hat{Y}_j$ denote the synthesized target sequence, and we use the L1 loss to impose a similarity measure between $\hat{Y}_j$ and the ground-truth Y_j. Therefore, we formulate our overall loss as:

$$\mathcal{L}_{overall} = \sum_{j} \mathcal{L}_1(\hat{Y}_j, Y_j) + \lambda \mathcal{L}_s, \tag{5}$$

where λ is the weighting parameter balancing the two terms.

3 Experiments

3.1 Data Acquisition

The imaging data were acquired on a 3T United Imaging uMR890 MRI scanner using a 64-channel head coil. Ten healthy subjects aged between 22 and 24 years, including 7 males and 3 females, were included in the study after providing informed consent.

Structural MRI with various contrasts was collected using a Fast Spin Echo (FSE) sequence, with adjustments made to imaging parameters such as repetition time (TR), echo time (TE), and inversion time (TI). For each subject, we gathered clinically popular structural MR images, including T1-FLAIR (denoted as “1” in Fig. 1 and Table 1), T2-weighted (“7”), T2-FLAIR (“9”), and Proton Density (PD) images (“5”). Additionally, we obtained further contrast images by sampling parameters around these four structural images.

Table 1. Imaging parameters of all sequences.

Seq	1	2	3	4	5	6	7	8	9
TR (ms)	2479	2479	6777	6777	8000	8000	8142	8000	8000
TE (ms)	9.62	10.42	10.42	62.52	10.42	39.92	104.2	99.98	98.78
TI (ms)	1055	0	0	0	0	0	0	1030	2025

The imaging protocol included 25 axial slices with a spacing of $0.64mm \times 0.64mm \times 6.5mm$ and a field of view (FOV) of $200mm \times 230mm$. For T1-FLAIR and T2-FLAIR sequences, the inversion time (TI) was empirically determined following the rule:

$$TI = T1[\ln 2 - \ln(1 + e^{-\frac{(TR - TE_{last})}{T1}})], \tag{6}$$

where TE_{last} is the last echo time. The detailed sequence information is shown in Table 1.

For preprocessing, we conducted bias field correction and affine registration within each subject using the Advanced Normalization Tools (ANTs) software package. All images are applied min-max normalization to scale the intensity range to $[0, 1]$.

3.2 Implementation Details

We implemented our method using the PyTorch backend, and the subsequent experiments were conducted on a server equipped with an Nvidia A100 GPU. We used the Adam optimizer, with the learning rate set to 10^{-4} and decayed by 0.5 every 200 epochs. The model was trained for 1000 epochs. All images were resized to 192×192 to accommodate the competing method. We utilized eight subjects as the training set and two subjects as the test set. The weighting parameter λ for the loss function in Eq. (5) was set to 0.1.

Table 2. Performance of different methods on the test dataset.

Model	PD to T2-FLAIR		PD to T1-FLAIR		PD to T2	
	PSNR	SSIM%	PSNR	SSIM%	PSNR	SSIM%
pGAN [5]	26.05±1.28	90.47±1.30	24.50±0.80	83.01±0.80	29.47±0.64	92.58±1.04
PTNet [18]	21.24±1.77	81.24±1.04	21.16±1.30	77.14±1.18	26.07±0.52	93.00±0.59
ResVIT [4]	26.12±1.11	91.72±1.95	25.04±0.60	87.72±0.60	28.38±0.57	96.63±0.42
Ours	**29.40±0.76**	**92.18±0.90**	**26.21±0.42**	**88.05±0.64**	**31.68±0.44**	**96.91±0.27**

Model	T2 to PD		T2 to T1-FLAIR		T2 to T2-FLAIR	
	PSNR	SSIM%	PSNR	SSIM%	PSNR	SSIM%
pGAN [5]	30.08±0.60	92.81±1.64	27.18±1.29	89.68±1.40	24.90±0.52	84.64±1.25
PTNet [18]	27.24±0.73	93.07±0.89	21.90±0.83	82.59±0.67	22.00±0.62	79.87±1.13
ResVIT [4]	29.48±0.81	96.51±1.60	27.23±1.48	92.54±1.59	24.41±0.52	87.96±1.45
Ours	**30.87±0.33**	**96.89±0.4**	**28.97±1.10**	**92.62±0.90**	**25.42±0.73**	**88.07±0.66**

3.3 Comparison with State-of-the-Art Methods

We compared our method with pGAN [5], PTNET [18], and ResViT [4] using 5-fold cross-validation on the dataset under six configurations of input-output combinations: “5” (PD) to “1” (T1-FLAIR), “5” (PD) to “9” (T2-FLAIR), “5” (PD) to “7” (T2), “7” (T2) to “5” (PD), “7” (T2) to “1” (T1-FLAIR), and “7” (T2) to “9” (T2-FLAIR). All models were trained from scratch.

It is noteworthy that for all competing methods, a separate model was trained for each respective task. In contrast, our approach involved training a single model for all tasks. We employed the commonly used Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) for quantitative evaluation of the synthesis quality. The quantitative results are shown in Table 2. Qualitative results are shown in Fig. 3. The rows from top to bottom illustrate the synthesized results from T2 to PD, T1-FLAIR, and T2-FLAIR, respectively. From the synthesized images, we observe that our method is most comparable to the ground truth, with fewer errors shown in the heatmaps. In summary, our method can better maintain the anatomical structure of the original image over other comparison models.

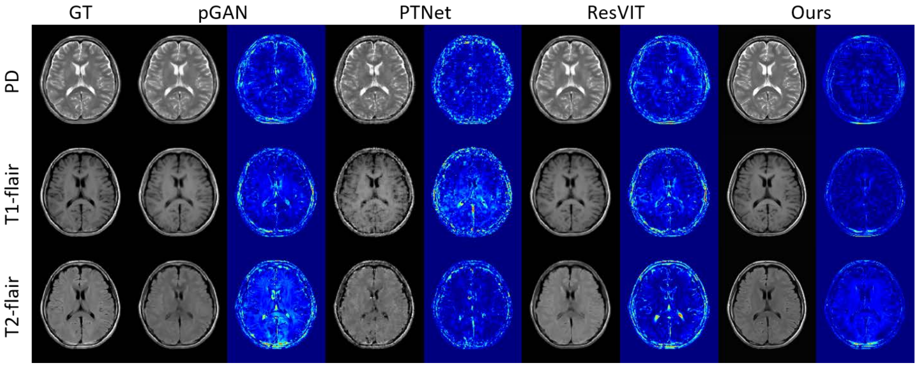

Fig. 3. Visualization of synthesized images and error maps of all methods.

3.4 Modulating Contrast across All Sequences

To validate the effectiveness of our method in modulating all contrasts, we evaluated our method with 5-fold cross-validation on the dataset under all configurations of input-output combinations. Quantitative results are shown in Table 3.

It is worth noting that during training, we only utilized sequences with indices [1, 2, 3, 5, 7, 8, 9], while sequences “4” and “6” were exclusively reserved for validation purposes. Qualitative results are shown in Fig. 4. We selected sequence “7” (T2) to exemplify the synthesis results of all other sequences. The first row in the figure shows the ground-truth MR images for all sequences. The second

Table 3. Performance of different input-output combinations in PSNR (dB). The rows indicate the input sequences and the columns indicate the target output sequences. The "*" indicates the sequences that are not seen during training.

Seq	1	2	3	4*	5	6*	7	8	9	Avg
1	-	28.18	24.89	24.45	24.75	19.74	24.45	24.89	24.75	24.08
2	29.54	-	26.01	26.43	24.98	19.45	27.08	22.37	24.47	24.52
3	27.64	27.65	-	27.88	30.75	22.99	28.80	24.42	24.64	25.58
4*	28.34	28.50	29.96	-	28.40	20.69	20.42	25.28	26.09	25.43
5	29.40	28.93	30.22	21.40	-	21.93	31.68	26.22	26.21	26.26
6*	23.71	19.64	20.92	22.37		-	19.48	17.83	19.37	20.64
7	28.97	28.43	32.90	29.23	30.87	25.97	-	26.65	25.42	27.77
8	33.41	22.04	27.74	20.54	20.21	20.00	35.89	-	24.52	25.60
9	29.98	26.33	24.89	25.17	25.29	22.12	24.56	24.50	-	24.86

row depicts the results obtained by using sequence "7" as input to synthesize images for the other sequences. The results demonstrated the generalizability of our approach, as the generated outcomes from sequence "7" remained consistent with the ground truth, regardless of whether they appeared in the training data or not.

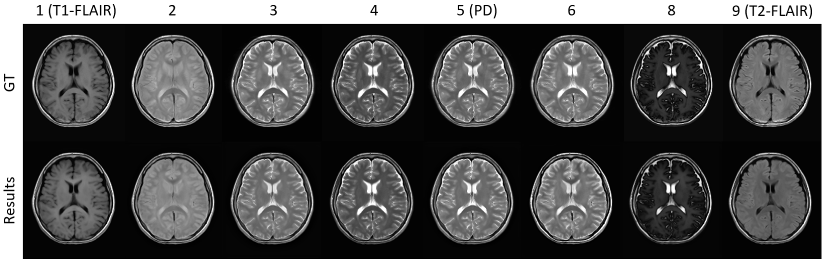

Fig. 4. Visualization of synthesized images using sequence "7" (T2) as input.

4 Conclusion

In this study, we introduced Contrast Representation Learning (CRL), an approach for MR image synthesis that transcends the conventional paradigm of direct cross-domain mapping. By exploring the high-dimensional manifold where MR images reside and considering the interplay of imaging parameters, CRL aims to capture the nuanced variations in contrast and enhance the fidelity of synthesized MR images. Notably, by constructing the contrast manifold based on imaging parameters, CRL demonstrates the capability to generate new contrast images, offering promising prospects for clinical applications.

Acknowledgments. This work was partially supported by National Natural Science Foundation of China (62131015) and Shanghai Municipal Central Guided Local Science and Technology Development Fund (YDZX20233100001001).

Disclosure of Interests. The authors declare no competing interests.

References

1. Alexander, D.C., Zikic, D., Zhang, J., Zhang, H., Criminisi, A.: Image quality transfer via random forest regression: applications in diffusion mri. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 225–232. Springer (2014)
2. Armanious, K., Jiang, C., Fischer, M., Küstner, T., Hepp, T., Nikolaou, K., Gatidis, S., Yang, B.: Medgan: Medical image translation using gans. Computerized medical imaging and graphics **79**, 101684 (2020)
3. Chartsias, A., Joyce, T., Giuffrida, M.V., Tsaftaris, S.A.: Multimodal mr synthesis via modality-invariant latent representation. IEEE transactions on medical imaging **37**(3), 803–814 (2017)
4. Dalmaz, O., Yurt, M., Çukur, T.: Resvit: residual vision transformers for multimodal medical image synthesis. IEEE Transactions on Medical Imaging **41**(10), 2598–2614 (2022)
5. Dar, S.U., Yurt, M., Karacan, L., Erdem, A., Erdem, E., Çukur, T.: Image synthesis in multi-contrast mri with conditional generative adversarial networks. IEEE Transactions on Medical Imaging **38**(10), 2375–2388 (2019)
6. Jog, A., Carass, A., Roy, S., Pham, D.L., Prince, J.L.: Random forest regression for magnetic resonance image synthesis. Medical image analysis **35**, 475–488 (2017)
7. Lan, H., Initiative, A.D.N., Toga, A.W., Sepehrband, F.: Sc-gan: 3d self-attention conditional gan with spectral normalization for multi-modal neuroimaging synthesis. BioRxiv pp. 2020–06 (2020)
8. Lee, D., Kim, J., Moon, W.J., Ye, J.C.: Collagan: Collaborative gan for missing image data imputation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 2487–2496 (2019)
9. Liu, J., Pasumarthi, S., Duffy, B., Gong, E., Datta, K., Zaharchuk, G.: One model to synthesize them all: Multi-contrast multi-scale transformer for missing data imputation. IEEE Transactions on Medical Imaging (2023)
10. Qin, Z., Liu, Z., Zhu, P., Ling, W.: Style transfer in conditional gans for cross-modality synthesis of brain magnetic resonance images. Computers in Biology and Medicine **148**, 105928 (2022)
11. Roy, S., Jog, A., Carass, A., Prince, J.L.: Atlas based intensity transformation of brain mr images. In: International Workshop on Multimodal Brain Image Analysis. pp. 51–62. Springer (2013)
12. Wang, G., Gong, E., Banerjee, S., Martin, D., Tong, E., Choi, J., Chen, H., Wintermark, M., Pauly, J.M., Zaharchuk, G.: Synthesize high-quality multi-contrast magnetic resonance imaging from multi-echo acquisition using multi-task deep generative model. IEEE transactions on medical imaging **39**(10), 3089–3099 (2020)
13. Wang, Z., Bovik, A., Sheikh, H., Simoncelli, E.: Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing **13**(4), 600–612 (2004)

14. Xin, B., Hu, Y., Zheng, Y., Liao, H.: Multi-modality generative adversarial networks with tumor consistency loss for brain mr image synthesis. In: 2020 IEEE 17th international symposium on biomedical imaging (ISBI). pp. 1803–1807. IEEE (2020)
15. Yang, H., Sun, J., Yang, L., Xu, Z.: A unified hyper-gan model for unpaired multi-contrast mr image translation. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part III 24. pp. 127–137. Springer (2021)
16. Yang, H., Lu, X., Wang, S.H., Lu, Z., Yao, J., Jiang, Y., Qian, P.: Synthesizing multi-contrast mr images via novel 3d conditional variational auto-encoding gan. Mobile Networks and Applications **26**, 415–424 (2021)
17. 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 transactions on medical imaging **38**(7), 1750–1762 (2019)
18. Zhang, X., He, X., Guo, J., Ettehadi, N., Aw, N., Semanek, D., Posner, J., Laine, A., Wang, Y.: Ptnet3d: A 3d high-resolution longitudinal infant brain mri synthesizer based on transformers. IEEE Transactions on Medical Imaging **41**(10), 2925–2940 (2022)