Hierarchical Anatomy-Aware Guidance for Brain Tissue Microstructure Reconstruction from T1-weighted MRI

Yuxing Li and Chuyang Ye(✉)

School of Integrated Circuits and Electronics, Beijing Institute of Technology, Beijing, China
chuyang.ye@bit.edu.cn

Abstract. Tissue microstructure information reconstructed from diffusion *magnetic resonance imaging* (MRI) provides crucial brain tissue information for brain disease analysis. However, clinical imaging time constraints often limit the availability of diffusion MRI, thus prompting research into tissue microstructure reconstruction from clinically feasible MRI modalities, such as T1-weighted MRI. Recent Transformer-based generative adversarial networks demonstrate potential by capturing long-range dependencies via self-attention in general MRI synthesis tasks, yet the significant gap between diffusion and T1-weighted MRI limits their ability to achieve optimal performance, leading to anatomical inconsistency in the reconstructed tissue microstructure maps. To address the problem, we propose a *hierarchical anatomy-aware guidance* (HAAG) framework for brain tissue microstructure reconstruction from T1-weighted MRI. First, we consider a two-level strategy to introduce the anatomical priors for the Transformer. At the input level of the Transformer, we propose an adaptive semantic embedding module that seamlessly integrates anatomical structure category information, providing semantic-level guidance for tissue microstructure reconstruction. At the feature modeling level of the Transformer, we propose a distance-guided self-attention mechanism to achieve effective information fusion of anatomical structures that balances both global and local contexts. Then, we consider a more general approach to verify the anatomical consistency at the output level of the whole synthesis network. We develop an anatomy-aware discriminative loss that encourages anatomical consistency between the input and output modalities. HAAG was validated on a public brain MRI dataset for reconstruction of tissue microstructure from T1-weighted MRI. The results demonstrate that our method significantly improves the quality of tissue microstructure reconstruction.

Keywords: Tissue Microstructure Reconstruction · Anatomical Prior · Consistency Generation.

Chuyang Ye is the corresponding author. Email: chuyang.ye@bit.edu.cn.

1 Introduction

Tissue microstructure information reconstructed from diffusion *magnetic resonance imaging* (MRI) has demonstrated great value into human brain researches [10]. However, the demand for long scanning times makes the acquisition of diffusion MRI impractical in clinical settings, thereby impeding the use of tissue microstructure information for clinical use. To address the problem, it is possible to apply generic MRI synthesis approaches so that tissue microstructure reconstruction is achieved via synthesis from clinically feasible MRI modalities, such as T1-weighted MRI. For example, earlier MRI synthesis works use registration-based strategies [15], compressed sensing [18], and random forest regression [9]. More Recent methods use deep learning techniques, mostly with generators and discriminators trained by adversarial learning [6,7,27]. These methods are based on *convolutional neural networks* (CNNs) [5,12,13,1,23,14] or Transformer-based generators [4,22]. Additionally, diffusion models [16,8,17] have been developed for better MRI synthesis performance, but they usually require considerable computational resources.

Among the DL models, the Transformer-based *generative adversarial network* (GAN) is more practical in resource-limited clinical scenarios with moderate computational efficiency. Thus, we focus on the use of Transformer-based MRI synthesis networks for brain tissue microstructure reconstruction from clinically feasible MRI. However, direct application of existing Transformer-based models for tissue microstructure reconstruction only leads to suboptimal performance due to the insensitivity of clinically feasible MRI to tissue microstructure, which greatly increases the synthesis difficulty and leads to anatomical inconsistency in the synthesized result. Thus, improvement of Transformer-based methods is still required for brain tissue microstructure reconstruction.

In this study, we introduce a *Hierarchical Anatomy-Aware Guidance* (HAAG) framework into Transformer-based brain MRI synthesis models for tissue microstructure reconstruction from clinically feasible MRI. HAAG uses anatomical prior knowledge, which has been neglected by previous synthesis methods, to guide the synthesis model at different levels. First, to improve the Transformer framework, a two-level approach is suggested for integrating the anatomical prior knowledge. At the input level, an adaptive semantic embedding module is proposed to integrate anatomical structure category information for semantic guidance. At the feature modeling level, a distance-guided self-attention mechanism is proposed to balance global and local contexts and blend information across anatomical structures. Second, to obtain a more general way to maintain anatomical consistency during network training, at the output level of the whole generator, an anatomy-aware discriminative loss is used to enforce anatomical agreement between input and output images, which encourages structural precision of the reconstruction results. For validation we applied the proposed methods to synthesize the NODDI [25] tissue microstructure maps, including the *Intra-Cellular Volume Fraction* (ICVF) and *Orientation Dispersion* (OD) maps, from T1-weighted MRI with the public *Human Connectome Project* (HCP) [20]

dataset. The results show that HAAG improves the quality of tissue microstructure reconstruction compared with other approaches.

2 Methods

2.1 Problem Formulation

Given a clinically feasible brain MRI modality, we aim to synthesize brain tissue microstructure maps from it. We assume that a training dataset of source MRI scans $\mathbf{X}^{\mathrm{s}}$ (i.e., clinically feasible MRI) and the corresponding target tissue microstructure maps $\mathbf{X}^{\mathrm{t}}$ from the same subjects are available. A synthesis model is trained on this paired data to learn the mapping from source to target images.

2.2 HAAG: Hierarchical Anatomy-Aware Guidance

We assume that the incorporation of prior knowledge about brain anatomy can aid tissue microstructure synthesis from the source MRI modality. To this end, we propose the HAAG framework. It combines two methods to integrate anatomical prior knowledge. First, we introduce the design of learnable semantic embedding with anatomical category information for the input of self-attention. Then, we explain how relative distance prior knowledge benefits self-attention. Second, we introduce a broader strategy for consistency-aligned feature generation and then present the overall brain MRI synthesis model with HAAG.

Adaptive Semantic Embedding, ASE In the original patch embedding, the input image feature $\mathbf{x} \in \mathbb{R}^{H \times W \times D_{\mathrm{token}}}$ was reshaped into a flattened 2D sequence of patches $\mathbf{x}_p \in \mathbb{R}^{N \times (P^2 \cdot C)}$, where H, W, C, N and D_{token} are respectively the height, the width, the feature channel of input, the number of patches and the embedding dimension of the Transformer tokens, P^2 is the patch size, and $N = HW/P^2$.

In the proposed *Adaptive Semantic Embedding* (ASE) module, the anatomical sequence for each patch which is denoted by $\mathbf{C}_p \in \{1, 2, \ldots, K\}^{P^2 \cdot C}$ with K as the number of anatomical categories, is firstly mapped to an embedding space using a learnable embedding matrix $\mathbf{E} \in \mathbb{R}^{K \times D_{\mathrm{emb}}}$ with embedding dimension D_{emb}:

$$\mathbf{Em}_p = \mathbf{E}[\mathbf{C}_p] \in \mathbb{R}^{N \times (P^2 \cdot D_{\mathrm{emb}})}. \tag{1}$$

Then, the embedding matrix $\mathbf{Em}_p$ is projected to the same feature space of the image feature tokens using a learnable linear transformation matrix $\mathbf{W}_{\mathrm{s}} \in \mathbb{R}^{(P^2 \cdot D_{\mathrm{emb}}) \times D_{\mathrm{token}}}$ to gather patch's semantic information from spatial perspective:

$$\mathbf{ASE}_p = \mathbf{Em}_p \cdot \mathbf{W}_{\mathrm{s}} \in \mathbb{R}^{N \times D_{\mathrm{token}}}. \tag{2}$$

Finally, the adaptive semantic embedding $\mathbf{ASE}_p$, image feature embedding $\mathbf{X}_p$, and positional embedding $\mathbf{PE}_p$ are fused as follows:

$$\mathbf{X}_p^{'} = \mathbf{X}_p + \mathbf{ASE}_p + \mathbf{PE}_p. \tag{3}$$

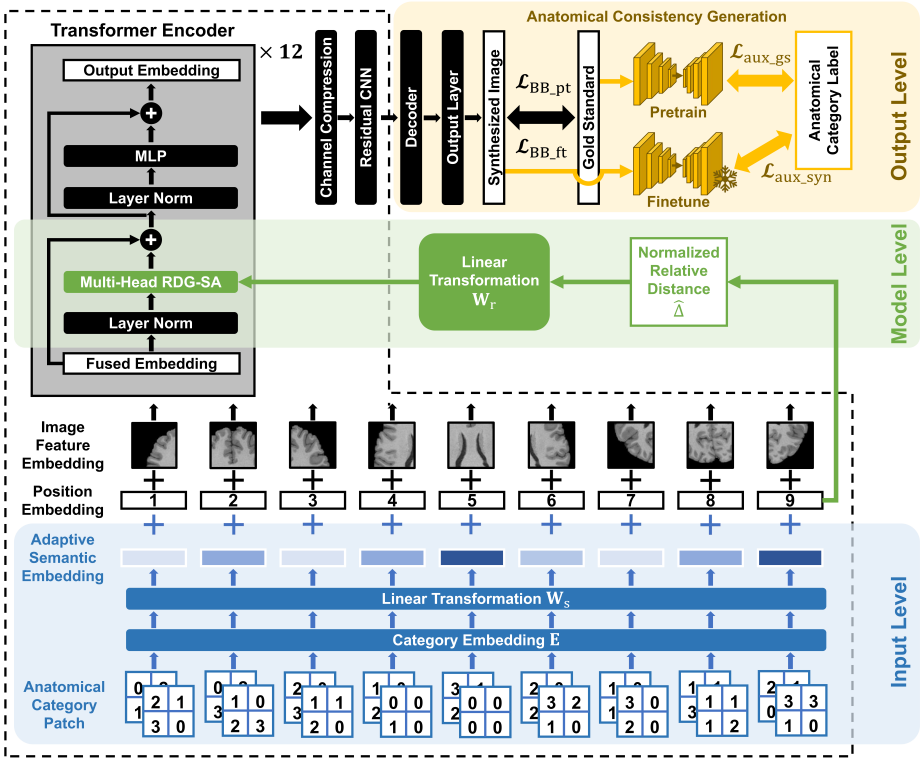

Fig. 1. A schematic of the generator of the brain MRI synthesis model for tissue microstructure reconstruction with Hierarchical Anatomy-Aware Guidance framework.

Note that to prevent overfitting, dropout is applied to the fused embedding $\mathbf{X}_p^{'}$ with a dropout probability $\gamma = 0.1$.

Relative Distance-Guided Self-Attention, RDG-SA Suppose N image tokens are sent into the self-attention module, and the i-th token is denoted by a row vector $\boldsymbol{X}_i^{'} \in \mathbb{R}^{D_{\text{token}}}$. For each token $\boldsymbol{X}_i^{'}$, the spatial position (x_i, y_i) can be obtained by:

$$x_i = \left\lfloor \frac{i-1}{N_W} \right\rfloor, \quad y_i = (i-1) \bmod N_W, \tag{4}$$

where $i \in \{1, 2, \ldots, N\}$ and $N_W = \frac{W}{P}$.

Given the spatial position, we have the horizontal and vertical relative distances for each pair of tokens $\boldsymbol{X}_i^{'}$ and $\boldsymbol{X}_j^{'}$ as:

$$\Delta x_{ij} = |x_i - x_j|, \quad \Delta y_{ij} = |y_i - y_j|. \tag{5}$$

Thus, the total relative distance Δ_{ij} is defined as the sum of the horizontal and vertical distances:

$$\Delta_{ij} = \Delta x_{ij} + \Delta y_{ij}. \tag{6}$$

Then, the relative distance is obtained by being normalized with the maximum possible distance $\Delta_{\max} = N_H + N_W$, where $N_H = \frac{H}{P}$:

$$\hat{\Delta}_{ij} = \frac{\Delta_{ij}}{\Delta_{\max}}. \tag{7}$$

With the spatial position prior knowledge about brain anatomy, it is possible to construct a relative distance-guided relation matrix $\mathbf{R} \in \mathbb{R}^{N \times N}$ for the input tokens, where the (i, j)-th entry $R_{i,j}$ of $\mathbf{R}$ roots from the complement of normalized relative distance $\hat{\Delta}_{ij}$, forecasting that higher $R_{i,j}$ indicates that i-th token and j-th token are more likely to inform each better:

$$\mathbf{R} = \mathbf{W}_{\mathrm{r}}(\mathbf{1} - \hat{\mathbf{\Delta}}), \tag{8}$$

where $\mathbf{W}_{\mathrm{r}} \in \mathbb{R}^{N \times N}$ is a learnable linear transformation matrix.

We propose to supplement the conventional self-attention with relation matrix $\mathbf{R}$ and design the *Relative Distance-Guided Self-Attention* (RDG-SA) as:

$$\text{RDG-SA}(\mathbf{X}^{'}) = \text{softmax}\left(\frac{\mathbf{X}^{'}\mathbf{W}_{\mathrm{Q}}(\mathbf{X}^{'}\mathbf{W}_{\mathrm{K}})^{\top}}{\sqrt{d_k}} + \mathbf{R}\right)\mathbf{X}^{'}\mathbf{W}_{\mathrm{V}}, \tag{9}$$

where $\mathbf{W}_{\mathrm{Q}}$, $\mathbf{W}_{\mathrm{K}}$, and $\mathbf{W}_{\mathrm{V}}$ are three learnable projection matrices in the Transformer, and d_k is the projected dimension. In this way, this enhanced attention mechanism prioritizes interactions between spatially proximate tokens, effectively modeling the local dependencies.

Multi-head self-attention plays a crucial role in the success of the Transformer [21]. RDG-SA can be extended to multi-head RDG-SA as well, which allows diverse attention for richer representations.

Anatomical Consistency Generation, ACG In order to explicitly ensure anatomical consistency between the synthesized and target images, we propose an *Anatomical Consistency Generation* (ACG) method where a perceptual loss is introduced.

First, a simple UNet-like perceptual network is constructed with 4 downsampling and 4 upsampling blocks to classify tissue types. Pretrained on gold standard images, it uses the loss function:

$$\mathcal{L}_{\mathrm{pt}} = \mathcal{L}_{\mathrm{BB\_pt}} + \mathcal{L}_{\mathrm{aux\_gs}}, \tag{10}$$

where $\mathcal{L}_{\mathrm{BB\_pt}}$ is the backbone (BB)'s pretraining loss and $\mathcal{L}_{\mathrm{aux\_gs}}$ is the cross-entropy loss computed between the gold standard images' classification result and tissue-category labels.

During the fine-tuning phase, the perceptual network is frozen, and the synthesized images are used as input to compute the tissue-type classification error as the loss to back-propagate:

$$\mathcal{L}_{\mathrm{ft}} = \mathcal{L}_{\mathrm{BB\_ft}} + \mathcal{L}_{\mathrm{aux\_syn}}, \tag{11}$$

where $\mathcal{L}_{\mathrm{BB\_ft}}$ is BB's fine-tuning loss and $\mathcal{L}_{\mathrm{aux\_syn}}$ is the cross-entropy loss computed between the synthesized images' classification result and the tissue-category labels.

Overall Synthesis Model Integrating ASE and RDG-SA forms the first component of the HAAG framework, concentrating on enhancing the Transformer by integrating anatomical priors. The second component of HAAG framework is the ACG module, verifying the anatomical consistency of generator's intermediate feature. The architectures of these three mechanisms are respectively shown in Fig. 1.

2.3 Implementation Details

Our selected Transformer-based model ResViT [4] has achieved superior performance in brain MRI synthesis. It has a generator and a discriminator, trained adversarially. Unlike CNN-based models that capture only local contexts, ResViT improves the generator's bottleneck with a Transformer-based module using self-attention for long-range dependencies.

The implementation of our method is based on the open-source code of ResViT. The data preprocessing process includes: (1) computing tissue microstructure and brain parcellation maps respectively from diffusion MRI and structural MRI, and (2) co-registering diffusion MRI and microstructure maps with structural MRI. Like ResViT [4], the proposed method takes 2D MRI slices as input, and the output slices are concatenated to form the 3D synthesis result. The training procedure also follows ResViT, which includes a pretraining phase and a fine-tuning phase. For more details that are not related to the HAAG framework, please refer to ResViT [4]. The codes for the proposed HAAG framework are available at https://github.com/Winterborner/HAAG.

3 Experiments

Data Description and Experimental Settings. To evaluate our proposed methods, we focused on synthesizing ICVF and OD maps from T1-weighted images. This is challenging because ICVF and OD, usually reconstructed from diffusion MRI [25], capture tissue microstructure while T1-weighted images aren't sensitive to it. We chose T1-weighted images as the source modality as they're commonly available in MRI acquisitions.

Experiments were performed on the publicly available HCP dataset [20]. We selected 61 subjects: 36 for training with 1 subject set aside for validation and 25 for testing. Each subject has a T1-weighted scan (0.7 mm isotropic voxel size) and a diffusion MRI scan (270 diffusion gradients on three b-shells: $b = 1000, 2000, 3000$ s/mm^2 and 1.25 mm isotropic voxel size).

ICVF and OD maps estimated from diffusion MRI scans using AMICO [3] were used for training the model with training subjects and as the gold standard for evaluation with test subjects. These maps were rigidly registered with T1-weighted images using the registration tool in ANTs [2]. For anatomical prior knowledge, we used FSL-FAST [26] to perform brain parcellation on T1-weighted images. For clarity in the results, the first and second components of the HAAG framework are denoted as 'Ours1' and 'Ours2' respectively.

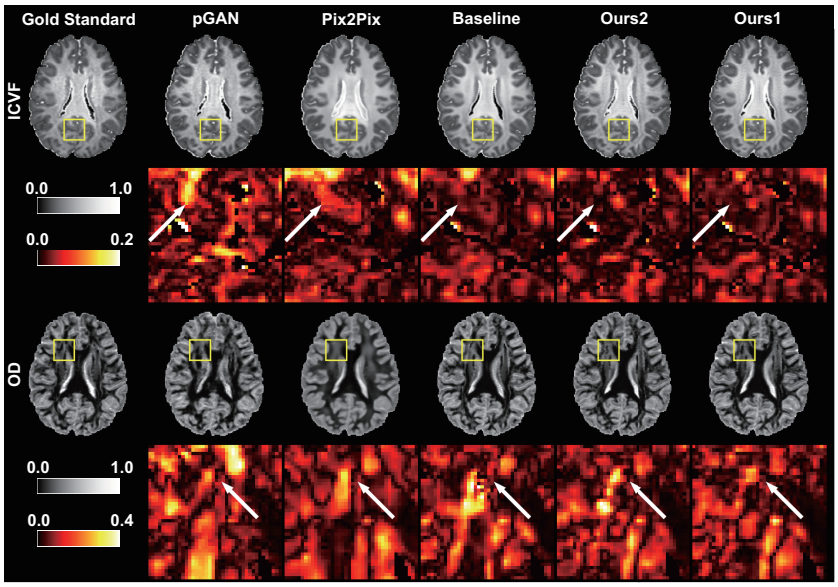

Fig. 2. Representative results of ICVF and OD synthesis and their zoomed error maps with gold standard shown for reference.

Evaluation of Synthesis Results We compared our approaches with three advanced image synthesis methods: pGAN [5], the 3D version of Pix2Pix (denoted by 'Pix2Pix') [19], and ResViT (denoted by 'Baseline') [4]. An example of the synthesized result of the proposed methods is shown in Fig. 2, where the gold standard is shown as well. Also, the zoomed error maps for the square regions highlighted in the synthesis results are displayed. The zoomed error maps of both proposed methods appear darker than those of competing methods, which indicates better agreement of the proposed methods with the gold standard. Moreover, the synthesized results of both proposed methods contain more accurate details than those of competing methods.

Quantitatively, we measured the *mean absolute error* (MAE), *root mean squared error* (RMSE), and *peak signal-to-noise ratio* (PSNR) of the synthesis results of the proposed and baseline methods, and presented results in Table 1. Ours1 had the lowest MAE and RMSE, along with the highest PSNR, while Ours2 was the second-best. These results show the superior synthesis quality of both proposed methods. Additionally, for all cases, the difference between the optimal method Ours1 and competing methods is statistically significant ($p < 0.001$) with a large effect size ($d > 0.8$).

Ablation Experiments The ablation study delved deeper into how ASE and RDG-SA affect the synthesis quality of Ours1, with results presented in Table 2. Note that the baseline backbone for this ablation study is that of 'Baseline'

Table 1. Means and *standard deviations* (stds) of the MAE, RMSE, and PSNR of the synthesis results for the proposed and competing methods. The best results are highlighted in bold. The proposed method with the best performance was compared with all competing methods with paired Student's t-tests ($^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$). The effect sizes (Cohen's d) for the comparison are also shown.

		MAE	RMSE	PSNR
		Mean±Std$^{p}_{d}$	Mean±Std$^{p}_{d}$	Mean±Std$^{p}_{d}$
ICVF	pGAN	$(5.83 \pm 0.29)_{\times 10^{-2}}{}^{***}_{4.06}$	$(8.53 \pm 0.38)_{\times 10^{-2}}{}^{***}_{4.50}$	$21.39 \pm 0.39^{***}_{4.24}$
	Pix2Pix	$(5.03 \pm 0.34)_{\times 10^{-2}}{}^{***}_{1.75}$	$(7.25 \pm 0.40)_{\times 10^{-2}}{}^{***}_{1.67}$	$22.81 \pm 0.47^{***}_{1.64}$
	Baseline	$(5.02 \pm 0.65)_{\times 10^{-2}}{}^{***}_{1.18}$	$(7.12 \pm 0.60)_{\times 10^{-2}}{}^{***}_{1.13}$	$22.98 \pm 0.73^{***}_{1.13}$
	Ours2	$(4.48 \pm 0.31)_{\times 10^{-2}}{}^{**}_{0.32}$	$(6.62 \pm 0.40)_{\times 10^{-2}}{}^{**}_{0.29}$	$23.60 \pm 0.54^{**}_{0.30}$
	Ours1	$\mathbf{(4.36 \pm 0.40) \times 10^{-2}}$	$\mathbf{(6.48 \pm 0.50) \times 10^{-2}}$	$\mathbf{23.79 \pm 0.68}$
OD	pGAN	$(9.86 \pm 0.15)_{\times 10^{-2}}{}^{***}_{16.91}$	$(12.70 \pm 0.18)_{\times 10^{-2}}{}^{***}_{16.28}$	$17.93 \pm 0.13^{***}_{15.04}$
	Pix2Pix	$(7.55 \pm 0.16)_{\times 10^{-2}}{}^{***}_{3.81}$	$(9.66 \pm 0.21)_{\times 10^{-2}}{}^{***}_{2.78}$	$20.31 \pm 0.19^{***}_{2.78}$
	Baseline	$(7.19 \pm 0.17)_{\times 10^{-2}}{}^{***}_{1.79}$	$(9.45 \pm 0.22)_{\times 10^{-2}}{}^{***}_{1.87}$	$20.50 \pm 0.20^{***}_{1.87}$
	Ours2	$(7.07 \pm 0.15)_{\times 10^{-2}}{}^{***}_{1.22}$	$(9.24 \pm 0.20)_{\times 10^{-2}}{}^{***}_{1.08}$	$20.69 \pm 0.19^{***}_{1.09}$
	Ours1	$\mathbf{(6.85 \pm 0.19) \times 10^{-2}}$	$\mathbf{(8.98 \pm 0.26) \times 10^{-2}}$	$\mathbf{20.93 \pm 0.25}$

Table 2. Means and stds of the MAE, RMSE, and PSNR from ablation experiments. The best results are highlighted in bold.

		MAE	RMSE	PSNR
		Mean±Std	Mean±Std	Mean±Std
ICVF	Ours1	$\mathbf{(4.36 \pm 0.40) \times 10^{-2}}$	$\mathbf{(6.48 \pm 0.50) \times 10^{-2}}$	$\mathbf{23.79 \pm 0.68}$
	w/o RDG-SA	$(4.49 \pm 0.34) \times 10^{-2}$	$(6.62 \pm 0.41) \times 10^{-2}$	23.60 ± 0.55
	w/o ASE	$(4.55 \pm 0.37) \times 10^{-2}$	$(6.77 \pm 0.43) \times 10^{-2}$	23.41 ± 0.55
OD	Ours1	$\mathbf{(6.85 \pm 0.19) \times 10^{-2}}$	$\mathbf{(8.98 \pm 0.26) \times 10^{-2}}$	$\mathbf{20.93 \pm 0.25}$
	w/o RDG-SA	$(6.91 \pm 0.14) \times 10^{-2}$	$(9.02 \pm 0.18) \times 10^{-2}$	20.89 ± 0.14
	w/o ASE	$(7.07 \pm 0.17) \times 10^{-2}$	$(9.25 \pm 0.22) \times 10^{-2}$	20.68 ± 0.21

in Table 1. ASE and RDG-SA were crucial in improving the precision of the synthesized images, where the complete setup resulted in optimal performance of Ours1.

4 Discussion and Conclusion

We have proposed the HAAG framework, two hierarchical anatomy-aware guidance methods, for brain tissue microstructure reconstruction from clinically feasible MRI. The HAAG framework, with its ASE-RDG-SA combination module and ACG module, incorporates brain anatomy prior knowledge at the Transformer's input and feature-modelling level, and the whole generator-output alignment level. This promotes anatomical consistency between source and target images. Experiments performed on the public HCP dataset for ICVF and OD

synthesis from T1-weighted images show that the HAAG framework improves the synthesis quality.

While recent related works have made significant contributions to structural preservation in image synthesis, notably mutual information loss for modality alignment [11] and hierarchical granularity approaches [24], our work represents the first framework leveraging structural MRI and anatomical priors specifically for brain tissue microstructure reconstruction. A limitation of our work is the restricted evaluation on pathological cases, which may affect model generalizability. In future work, we will validate the clinical utility of our method on diverse patient cohorts.

Acknowledgments. This work was supported by the Beijing Municipal Natural Science Foundation (7242273) and the Key Program of National Natural Science Foundation of China (82330057).

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

References

1. 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)
2. Avants, B.B., Tustison, N.J., Song, G., Cook, P.A., Klein, A., Gee, J.C.: A reproducible evaluation of ants similarity metric performance in brain image registration. Neuroimage **54**(3), 2033–2044 (2011)
3. Daducci, A., Canales-Rodríguez, E.J., Zhang, H., Dyrby, T.B., Alexander, D.C., Thiran, J.P.: Accelerated microstructure imaging via convex optimization (AMICO) from diffusion MRI data. NeuroImage **105**, 32–44 (2015)
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., Cukur, T.: Image synthesis in multi-contrast MRI with conditional generative adversarial networks. IEEE Transactions on Medical Imaging **38**(10), 2375–2388 (2019)
6. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial networks. Communications of the ACM **63**(11), 139–144 (2020)
7. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1125–1134 (2017)
8. Jiang, L., Mao, Y., Wang, X., Chen, X., Li, C.: Cola-diff: Conditional latent diffusion model for multi-modal mri synthesis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 398–408. Springer (2023)
9. 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)

10. Kamiya, K., Hori, M., Aoki, S.: Noddi in clinical research. Journal of neuroscience methods **346**, 108908 (2020)
11. Kang, M., Chikontwe, P., Won, D., Luna, M., Park, S.H.: Structure-preserving image translation for multi-source medical image domain adaptation. Pattern Recognition **144**, 109840 (2023)
12. 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)
13. Li, H., Paetzold, J.C., Sekuboyina, A., Kofler, F., Zhang, J., Kirschke, J.S., Wiestler, B., Menze, B.: DiamondGAN: unified multi-modal generative adversarial networks for MRI sequences synthesis. In: International Conference on Medical Image Computing and Computer Assisted Intervention. pp. 795–803. Springer (2019)
14. 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 **42**(9), 2577–2591 (2023)
15. Miller, M.I., Christensen, G.E., Amit, Y., Grenander, U.: Mathematical textbook of deformable neuroanatomies. Proceedings of the National Academy of Sciences **90**(24), 11944–11948 (1993)
16. Özbey, M., Dalmaz, O., Dar, S.U., Bedel, H.A., Özturk, Ş., Güngör, A., Çukur, T.: Unsupervised medical image translation with adversarial diffusion models. IEEE Transactions on Medical Imaging (2023)
17. Pan, S., Chang, C.W., Peng, J., Zhang, J., Qiu, R.L., Wang, T., Roper, J., Liu, T., Mao, H., Yang, X.: Cycle-guided denoising diffusion probability model for 3d cross-modality mri synthesis. arXiv preprint arXiv:2305.00042 (2023)
18. Roy, S., Carass, A., Prince, J.: A compressed sensing approach for mr tissue contrast synthesis. In: International Conference on Information Processing in Medical Imaging. pp. 371–383. Springer (2011)
19. Tang, Y., Yang, D., Li, W., Roth, H.R., Landman, B., Xu, D., Nath, V., Hatamizadeh, A.: Self-supervised pre-training of swin transformers for 3d medical image analysis. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 20730–20740 (2022)
20. Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E.J., Yacoub, E., Ugurbil, K.: The WU-Minn Human Connectome Project: An overview. NeuroImage **80**, 62–79 (2013)
21. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in Neural Information Processing Systems **30** (2017)
22. Yan, S., Wang, C., Chen, W., Lyu, J.: Swin transformer-based gan for multi-modal medical image translation. Frontiers in Oncology **12**, 942511 (2022)
23. 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)
24. Yu, Z., Zhao, B., Zhang, S., Chen, X., Yan, F., Feng, J., Peng, T., Zhang, X.Y.: Hifi-syn: Hierarchical granularity discrimination for high-fidelity synthesis of mr images with structure preservation. Medical Image Analysis **100**, 103390 (2025)
25. Zhang, H., Schneider, T., Wheeler-Kingshott, C.A., Alexander, D.C.: NODDI: practical in vivo neurite orientation dispersion and density imaging of the human brain. NeuroImage **61**(4), 1000–1016 (2012)

26. Zhang, Y., Brady, M., Smith, S.: Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm. IEEE transactions on medical imaging **20**(1), 45–57 (2002)
27. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2223–2232 (2017)