DPMix-GAN: Dual-domain Perception and Dynamic Patch-wise Mixing GAN for 3D CT Synthesis in MR-only Radiotherapy

Jiapeng Li[1,2], Mingzhu Li[1,2], Xiao Zhang[1,2], Shuyuan Zhang[3,4], Cheng Yuan[3,4], Lian Zhang[1,2,5(✉)], and Hongying Feng[6(✉)]

[1] Medical AI Lab, The First Hospital of Hebei Medical University, Hebei Medical University, Shijiazhuang, China
[2] Hebei Provincial Engineering Research Center for AI-Based Cancer Treatment Decision-Making, The First Hospital of Hebei Medical University, Hebei Medical University, Shijiazhuang, China
[3] Department of Oncology, Yichang Central People's Hospital, Yichang, China
[4] The First College of Clinical Medical Science, Yichang, China Three Gorges University, Yichang, China
[5] Department of Oncology, The First Hospital of Hebei Medical University, Hebei Medical University, Shijiazhuang, China
[6] College of Mathematics and Physics, China Three Gorges University, Yichang, China
lianzhang@hebmu.edu.cn; tracyfhy2015@outlook.com

Abstract. Magnetic resonance (MR) imaging provides excellent soft-tissue contrast without ionizing radiation, but lacks electron density information essential for radiotherapy dose calculation. Existing MR-to-CT synthesis methods mostly only depend on spatial-domain learning, leading to texture overfitting and compromised bone delineation and dosimetric accuracy. Critically, most latest deep learning approaches lack rigorous clinical validation on dose computation and treatment planning, limiting their translational utility. To address these gaps, we propose DPMix-GAN, a generative adversarial framework integrating dual-domain perception and patch-wise feature dynamic mixing for robust CT synthesis. DPMix-GAN fuses spatial and frequency features to capture complementary anatomical representations, and employs dynamic patch-wise mixing of high-order statistics to preserve anatomical structure. Experimental results on a clinical dataset across brain and pelvic regions shows that DPMix-GAN generated anatomically accurate synthesis CT with well-preserved boundaries. Moreover, clinical validation on radiotherapy treatment plans confirmed dosimetric equivalence between synthetic and real CT, with gamma passing rates exceeding clinical acceptance thresholds. The code is available at https://github.com/AprilT0621/DPMix-GAN.

Keywords: MR-to-CT synthesis · Generative Adversarial Networks · Transformer.

1 Introduction

Radiotherapy is a widely used treatment modality for cancer, in which radiation dose is delivered to target tissues while sparing surrounding normal organs. Computed Tomography (CT) has served as the primary imaging modality for radiotherapy treatment planning (RTP), as it directly provides the electron density maps required for accurate radiation dose computation. However, CT suffers from inferior soft-tissue contrast, which limits its effectiveness in precisely delineating tumor boundaries in anatomically complex regions. Magnetic resonance (MR) imaging has become increasingly important in modern RTP, primarily due to its excellent soft-tissue contrast and the absence of ionizing radiation [21] [1]. These advantages make MR particularly valuable for target delineation, especially in anatomically complex regions. However, MR images do not provide electron density information, which is essential for accurate dose calculation. As a result, reliable MR-to-CT synthesis is a key component for enabling MR-only RTP workflows [11]. In clinical practice, even subtle structural inaccuracies can propagate into significant dosimetric errors, highlighting the need for strict anatomical fidelity in synthetic CT (sCT) generation [13] [9].

Recent years have witnessed significant progress in deep learning–based MR-to-CT synthesis [12] [2] [3], among which diffusion models [5] have recently been explored as a promising technology for improving synthesis quality and structural realism. However, their application [16] [4] in RTP remains constrained by long inference times and the potential risk of stochastic anatomical variations, which are undesirable in dose-sensitive clinical settings. In contrast, Generative adversarial network (GAN)-based frameworks retain clear advantages in terms of inference efficiency and deterministic generation. Many current GAN-based approaches [19] primarily operate in the spatial domain, which limits their ability to model high-frequency anatomical details. Although recent studies have explored frequency-aware MR-to-CT synthesis by incorporating spectral representations into image generation [10], frequency information is typically used as an auxiliary representation and is not explicitly integrated into hierarchical feature learning. This often leads to over-smoothed bone structures and inaccurate representation of sharp tissue interfaces. At the same time, under the limited data regimes common in medical imaging, GANs are prone to texture overfitting. In such cases, discriminators tend to focus on modality-specific appearance cues rather than underlying anatomical structures, resulting in poor generalization across centers and patient cohorts. Moreover, most existing studies [8] [18] rely heavily on image-level similarity metrics, which do not necessarily correlate with dosimetric accuracy, leaving their practical value for RTP insufficiently validated.

This paper proposes an anatomy-driven MR-to-CT synthesis framework called DPMix-GAN. It aims to enhance anatomical perception during synthesis while suppressing modality-specific distractions during adversarial learning, thereby achieving reliable synthesis of sharp tissue that are crucial for radiotherapy dose calculation.

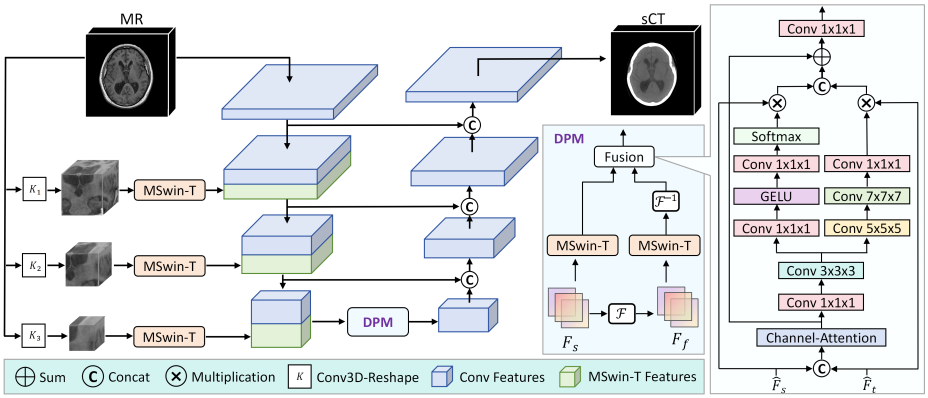

Fig. 1. Overview of the generator of DPMix-GAN.

Our main contributions are summarized as follows. **1)** We propose a Dual-Domain Perception Module (DPM) that jointly captures spatial and frequency-domain features to better preserve high-frequency anatomical boundaries in MR-to-CT synthesis. **2)** We introduce a 3D Patch-wise Feature Dynamic Mixing (PFDM) strategy to suppress modality-specific textures and promote anatomy-oriented representations. **3)** We conduct comprehensive image level and clinical level evaluations on multi-region datasets. The results demonstrate that the sCT images achieve strict clinical equivalence with real CT in treatment planning systems, supporting reliable MR-only radiotherapy.

2 Method

Framework Overview. DPMix-GAN is designed to achieve high-accuracy cross-modality MR-to-CT synthesis, which consists of a generator and a discriminator. The overview of the generator is illustrated in Fig. 1, which utilizes Multi-scale image tokens and Multi-shape Window Self-attention (MSwin-T) blocks to enlarge receptive fields for learning multi-directional spatial representations inspired by [22]. Our proposed Dual-Domain Perception Module (DPM) is embedded at the bottleneck to jointly model spatial features and frequency-domain responses, facilitating the preservation of fine-grained anatomical boundaries. The discriminator adopts a 3D PatchGAN [6] integrated with a Patch-wise Feature Dynamic Mixing (PFDM) module, which perturbs localized feature statistics to guide the adversarial in learning the consistency of anatomical structure.

Let $\{(x^i_{MR}, x^i_{CT})\}_{i=1}^{N}$ denote a set of paired training samples from the MR domain X and the corresponding ground-truth CT domain Y. Our DPMix-GAN framework aims to learn a cross-modal mapping $G : X \rightarrow Y$ that synthesizes high-fidelity CT volumes. To ensure that the synthesized images strictly follow the distribution of real clinical data, G is trained to fool the discriminator D through a min-max game. The discriminator D is optimized to distinguish real

MR-CT pairs from synthetic ones, while G strives to minimize this objective. The adversarial loss is defined as:

$$L_{adv} = \mathbb{E}_{x_{MR},x_{CT}} \left[\log D\left(x_{MR}, x_{CT}\right)\right] + \mathbb{E}_{x_{MR}} \left[\log\left(1 - D\left(x_{MR}, G\left(x_{MR}\right)\right)\right)\right] \quad (1)$$

Dual-domain Perception Module. After a series of downsampling operations, the MR image is transformed into spatial-domain features F_s. DPM applies a fast Fourier transform (FFT) to map F_s into the frequency domain:

$$F_f = \mathcal{F}(F_s) \quad (2)$$

where $\mathcal{F}$ denotes the FFT operation and F_f represents the corresponding frequency-domain features. The spatial features F_s and frequency features F_f are then independently fed into the MSwin-T to capture domain-specific global dependencies. Subsequently, the enhanced frequency-domain features are transformed back to the spatial domain via the inverse FFT:

$$\hat{F}_s = \text{MSwin-T}(F_s), \quad \hat{F}_f = \mathcal{F}^{-1}(\text{MSwin-T}(F_f)) \quad (3)$$

where $\mathcal{F}^{-1}$ denotes the inverse FFT operation. The Fusion block is designed to adaptively integrate $\hat{F}_s$ and $\hat{F}_f$, so as to exploit complementary information across domains and enhance structural representation:

$$F_{\text{out}} = \text{Fusion}(\hat{F}_s, \hat{F}_f), \quad (4)$$

where F_{out} denotes the fused output feature for subsequent decoding.

To minimize intensity differences and ensure accurate HU estimation, we implement the pixel-wise L1 distance loss L_{L1} to provide direct voxel-level supervision between the synthetic and real CT images:

$$L_{L1} = \|G(x_{MR}) - x_{CT}\|_1 \quad (5)$$

To further enhance the capture of fine textures and structural details, we incorporate a VGG-based perceptual loss L_{VGG}. Unlike pixel-wise losses, this term compares high-level feature maps extracted by a pre-trained VGG-19 network ϕ [14]. L_{VGG} is defined as:

$$L_{VGG} = \|\phi(G(x_{MR})) - \phi(x_{CT})\|_1 \quad (6)$$

3D Patch-wise Feature Dynamic Mixing. To prevent the discriminator from overfitting to modality-specific superficial textures, the 3D PFDM module partitions feature maps of the discriminator into localized volumetric patches, as illustrated in Fig.2. For a generated content feature patch x and a corresponding reference style patch y, PFDM computes four local high-order statistics: mean μ, variance σ^2, skewness γ, and kurtosis κ. Specifically:

$$\begin{gathered} \mu_x = \frac{1}{N}\sum x_i, \quad \sigma_x^2 = \frac{1}{N}\sum \left(x_i - \mu_x\right)^2 \\ \gamma_x = \frac{\frac{1}{N}\sum \left(x_i - \mu_x\right)^3}{\left(\sigma_x^2 + \epsilon\right)^{3/2}}, \quad \kappa_x = \frac{\frac{1}{N}\sum \left(x_i - \mu_x\right)^4}{\left(\sigma_x^2 + \epsilon\right)^2} \end{gathered} \quad (7)$$

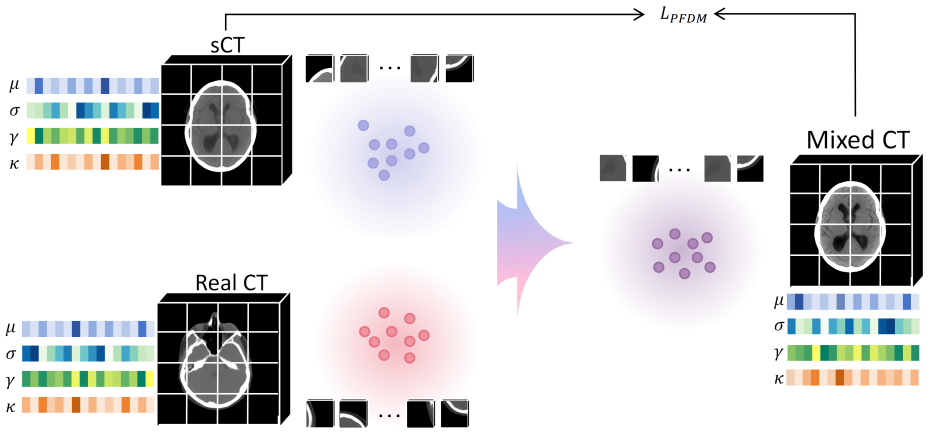

Fig. 2. Overview of PFDM.

After normalizing the content feature as $x_{norm} = (x - \mu_x)/\sqrt{\sigma_x^2 + \epsilon}$, the style statistics of y are injected to synthesize a perturbed feature:

$$x_{PFDM} = x_{norm}\sqrt{\sigma_y^2 + \epsilon} + \mu_y + (\gamma_y - \gamma_x)\, x_{norm}^3 \sqrt{\sigma_y^2 + \epsilon} + (\kappa_y - \kappa_x)\, x_{norm}^4 \sqrt{\sigma_y^2 + \epsilon} \tag{8}$$

Using a dynamically sampled parameter $\alpha \sim \mathcal{U}(0, 1)$, the features are mixed as follows to disrupt texture reliance:

$$x_{\text{mix}} = \alpha x + (1 - \alpha) x_{\text{PFDM}} \tag{9}$$

These patch-wise mixed features x_{mix} are spatially reassembled after every 3D convolutional layer during the discriminator, and the final output is denoted as $D_{PFDM}(G(x_{MR}), x_{CT})$. To actualize this structural regularization, we apply a consistency loss L_{PFDM}:

$$L_{PFDM} = \mathbb{E}_{x_{MR}, x_{CT}} \left[\| D\left(G(x_{MR})\right) - D_{PFDM}\left(G(x_{MR}), x_{CT}\right) \|_2^2 \right] \tag{10}$$

where $G(x_{MR})$ denote the sCT volume and x_{CT} represent a randomly sampled real CT reference from the training set.

This design reduces the sensitivity of the discriminator to localized style variations and encourages it to focus on cross-modal anatomical structures. The overall objective is defined as:

$$L_{\text{total}} = L_{\text{adv}} + \lambda_1 L_1 + \lambda_{\text{VGG}} L_{\text{VGG}} + \lambda_{\text{PFDM}} L_{\text{PFDM}} \tag{11}$$

where λ_{L1}, λ_{VGG} and λ_{PFDM} are hyper-parameters. This comprehensive loss strategy enables DPMix-GAN to generate synthetic CT images with high structural fidelity and realistic intensity distributions.

3 Experimental Results

Dataset and Preprocessing. A total of 193 paired CT–MR scans were retrospectively collected from a single medical institution, including brain (102 cases) and pelvic (91 cases) regions. T2-weighted sequences were acquired for brain MR images ,while T1-weighted sequences for pelvic MR images. The interval between CT and MR scans for each patient was within 7 days to minimize anatomical variation. This retrospective study was approved by the institutional review board (IRB, 13-005709). The brain data was split into 81/5/16 cases for training, validation, and testing, and the pelvic data into 72/4/15 cases, respectively.

For data preprocessing, we first applied N4 Bias Field Correction [17] to all MR images to remove intensity inhomogeneities. The images were then resampled to a resolution of $1.0 \times 1.0 \times 1.0$ mm^3 for brain data and $1.0 \times 1.0 \times 2.5$ mm^3 for pelvis data. Each CT image was aligned to the corresponding MR image using rigid registration [7]. All rigidly registered pairs were carefully inspected with marginal cases of localized mismatches observed and manually adjusted. Following the registration process, the brain volumes were cropped to preserve cranial anatomy. CT images were first clipped to the range of $[-1000, 2000]$ HU. Both MR and CT images were uniformly normalized to an intensity range of $[-1, 1]$. Finally, 3D patches with a size of $128 \times 128 \times 8$ were extracted for model training and evaluation.

Implementation Details and Evaluation Metrics. The proposed DPMix-GAN was implemented in Python using the PyTorch library. We trained DPMix-GAN for 200 epochs with an initial learning rate of 0.0002 that was dynamically adjusted using a "poly" decay strategy. We set the batch size to 4 and assigned $\lambda_{L1} = 20$, $\lambda_{VGG} = 1$ and $\lambda_{PFDM} = 0.01$ in the total loss function. All experiments were conducted on a NVIDIA RTX 4070S GPU with 12GB memory.

To quantitatively evaluate the quality of the synthetic CT images, we employed three widely accepted metrics, including the Mean Absolute Error (MAE), Structural Similarity Index Measure (SSIM), and the Peak Signal-to-Noise Ratio (PSNR). We further evaluated the translational value of DPMix-GAN through dosimetric validation besides conventional image-level metrics. Treatment plans

Table 1. Quantitative comparison on brain and pelvis regions.

Methods	Brain			Pelvis		
	MAE↓	SSIM↑	PSNR↑	MAE↓	SSIM↑	PSNR↑
3D-UNet	$37.3_{\pm 8.6}$	$0.892_{\pm 0.018}$	$26.9_{\pm 1.1}$	$48.9_{\pm 2.3}$	$0.793_{\pm 0.019}$	$29.9_{\pm 0.5}$
SwinUNETR	$27.3_{\pm 5.6}$	$0.920_{\pm 0.016}$	$28.7_{\pm 1.2}$	$32.2_{\pm 3.3}$	$0.878_{\pm 0.008}$	$\mathbf{30.2}_{\pm 1.3}$
MTT-Net	$30.3_{\pm 4.5}$	$0.908_{\pm 0.013}$	$28.0_{\pm 0.7}$	$46.9_{\pm 2.5}$	$0.739_{\pm 0.021}$	$28.3_{\pm 0.9}$
DPMix-GAN	$\mathbf{24.7}_{\pm 3.8}$	$\mathbf{0.927}_{\pm 0.011}$	$\mathbf{29.3}_{\pm 0.8}$	$\mathbf{30.7}_{\pm 3.4}$	$\mathbf{0.882}_{\pm 0.009}$	$\mathbf{30.2}_{\pm 1.2}$

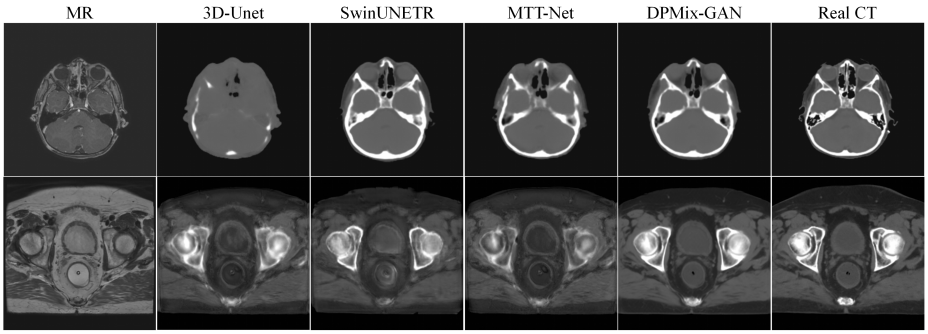

Fig. 3. Visualization of the synthesis results of different methods on brain and pelvis regions.

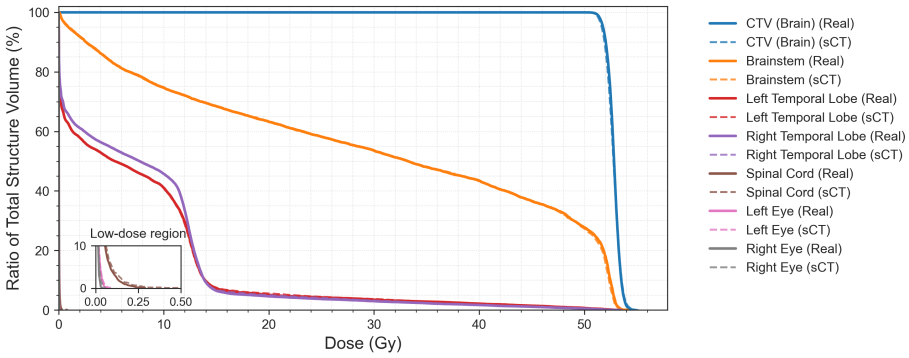

Fig. 4. DVH comparison between the sCT-based plan and the real CT-based plan for a representative patient.

optimized on the real CT were recalculated on the generated sCT using the MC-square Monte Carlo dose calculation algorithm implemented in Eclipse v15.6 to assess dosimetric equivalence. To quantitatively assess the 3D dose distribution consistency, Gamma index analysis is also performed.

Comparisons with Other Methods. The quantitative results in Table 1 demonstrate that DPMix-GAN consistently achieves the best performance across all evaluation metrics in both brain and pelvis regions. Specifically, it reduces the MAE by 2.60 HU and 1.45 HU in the brain and pelvis compared to SwinUNETR [15], and yields a MAE reduction of 16.18 HU in the pelvis over the multi-region designed MTT-Net [22]. The visualization is shown in Fig. 3. 3D-UNet [20] and MTT-Net produce over-smoothed images failing to delineate sharp anatomical interfaces and SwinUNETR introduces noticeable texture artifacts. In contrast, DPMix-GAN accurately reconstructs high-frequency details.

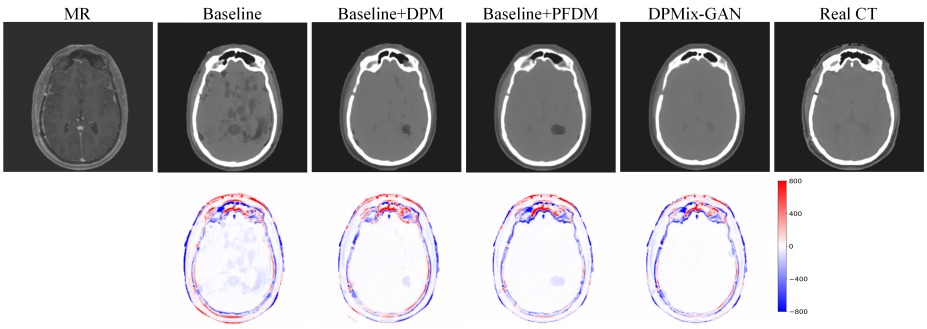

Fig. 5. Visualization of the synthesis results of ablation study on brain regions.

Table 2. Ablation study on brain and pelvis regions.

Methods	Brain			Pelvis		
	MAE↓	SSIM↑	PSNR↑	MAE↓	SSIM↑	PSNR↑
Baseline	$32.0_{\pm 5.5}$	$0.906_{\pm 0.017}$	$27.5_{\pm 0.9}$	$43.5_{\pm 2.8}$	$0.804_{\pm 0.015}$	$28.7_{\pm 0.8}$
Baseline+DPM	$27.2_{\pm 4.4}$	$0.921_{\pm 0.132}$	$28.6_{\pm 0.8}$	$36.7_{\pm 3.3}$	$0.846_{\pm 0.013}$	$29.5_{\pm 1.0}$
Baseline+PFDM	$26.7_{\pm 4.4}$	$0.921_{\pm 0.013}$	$28.8_{\pm 0.9}$	$36.0_{\pm 3.4}$	$0.842_{\pm 0.012}$	$30.0_{\pm 1.2}$
DPMix-GAN	$\mathbf{24.7}_{\pm 3.8}$	$\mathbf{0.927}_{\pm 0.011}$	$\mathbf{29.3}_{\pm 0.8}$	$\mathbf{30.7}_{\pm 3.4}$	$\mathbf{0.882}_{\pm 0.009}$	$\mathbf{30.2}_{\pm 1.2}$

Clinical Dosimetric Validation. As illustrated in Fig. 4, the Dose-Volume Histogram (DVH) curves for both the Clinical Target Volume (CTV) and critical Organs at Risk (OARs)—including the brainstem, temporal lobes, spinal cord, and eyes—exhibit near-perfect congruence between the real and sCTs. Our synthetic CT achieved a dose Gamma passing rate of 96.23% under the 2%/2mm criterion. These results exceed the standard clinical acceptance threshold, demonstrating that DPMix-GAN yields highly reliable electron density maps, enabling safe and accurate dose calculation for MR-only radiotherapy.

Ablation Study. We adopted a multi-scale 3D-UNet with MSwin-T and a standard 3D PatchGAN discriminator as baseline. Table 2 shows that incorporating either DPM or PFDM module leads to clear performance gains over the baseline, reducing the brain MAE from 32.0 HU to 27.2 HU and 26.7 HU, respectively. DPM enhances cross-domain feature representation by jointly modeling spatial and frequency information, while PFDM regularizes the discriminator by mitigating modality-specific texture sensitivity during adversarial learning. Notably, integrating both modules achieves the best result with an MAE of 24.7 HU, which is further supported by the visual error maps in Fig. 5. The baseline exhibits pronounced errors near skull boundaries and internal artifacts, whereas DPM improves boundary reconstruction and PFDM suppresses structural hallucinations, demonstrating their complementary effects.

4 Conclusion

In this paper, we introduce DPMix-GAN for high-fidelity MR-to-CT synthesis. The experimental results on brain and pelvis regions demonstrate that our model can effectively extract high-frequency anatomical boundaries and fuse them with spatial representations while suppressing texture overfitting, significantly improving structural fidelity and dose calculation accuracy. Further clinical dosimetric validation demonstrates that our synthetic CTs achieve strict equivalence with real CTs, yielding dose Gamma passing rates that exceed clinical acceptance thresholds. DPMix-GAN provides reliable electron density maps for MR-only radiotherapy, thereby supporting optimized treatment workflows. Future work will focus on validating the proposed method on public datasets and multi-center cohorts to further assess its generalizability across diverse imaging protocols and patient populations.

Acknowledgments. This work was supported by the Hebei Provincial Health Commission Medical Research Program (20250011, 20260098 and 20260115); the Innovation & Development Medical Cooperation Program of Hengrui-Hebei (HR202502085); the Hebei Provincial "Yanzhao Golden Terrace Talent Recruitment Program"- Top-Tier Talent Program (HY2025050008); the Chronic Disease Management Research Program of the Center for Capacity Building and Continuing Education, National Health Commission (GWJJMB202510022202); the National Foreign Experts Program (S20250235); the National Natural Science Foundation of China (12305393); the National Natural Science Foundation of Hebei Province (H2026206072); the "Spark" Scientific Research Program of the First Affiliated Hospital of Hebei Medical University (XH202515).

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

References

1. Computed tomography synthesis from magnetic resonance imaging using cycle generative adversarial networks with multicenter learning. Physics and Imaging in Radiation Oncology **28**, 100511 (2023)
2. 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)
3. Emami, H., Dong, M., Nejad-Davarani, S.P., Glide-Hurst, C.K.: Sa-gan: Structure-aware gan for organ-preserving synthetic ct generation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 471–481. Springer (2021)
4. Gong, C., Huang, Z., Wu, Z., Zhao, M., Xu, F., Bai, X., Feng, M., Granados, A., Meng, G., Lei, Z., Liu, H.: Mr-to-ct translation using frequency-separated diffusion models. In: 2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI). pp. 1–5 (2025)
5. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)

6. 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)
7. Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M.: Fsl. Neuroimage **62**(2), 782–790 (2012)
8. Li, Y., Xu, S., Chen, H., Sun, Y., Bian, J., Guo, S., Lu, Y., Qi, Z.: Ct synthesis from multi-sequence mri using adaptive fusion network. Computers in Biology and Medicine **157**, 106738 (2023)
9. Liang, X., Yen, A., Bai, T., Godley, A., Shen, C., Wu, J., Meng, B., Lin, M.H., Medin, P., Yan, Y., et al.: Bony structure enhanced synthetic ct generation using dixon sequences for pelvis mr-only radiotherapy. Medical physics **50**(12), 7368–7382 (2023)
10. Morell-Ortega, S., Ruiz-Perez, M., Gadea, M., Vivo-Hernando, R., Rubio, G., Aparici, F., de la Iglesia-Vaya, M., Tourdias, T., Mansencal, B., Coupé, P., et al.: Robust deep mri contrast synthesis using a prior-based and task-oriented 3d network. Imaging Neuroscience **3**, IMAG–a (2025)
11. Otazo, R., Lambin, P., Pignol, J.P., Ladd, M.E., Schlemmer, H.P., Baumann, M., Hricak, H.: Mri-guided radiation therapy: an emerging paradigm in adaptive radiation oncology. Radiology **298**(2), 248–260 (2021)
12. Phan, V.M.H., Xie, Y., Zhang, B., Qi, Y., Liao, Z., Perperidis, A., Phung, S.L., Verjans, J.W., To, M.S.: Structural attention: Rethinking transformer for unpaired medical image synthesis. In: International conference on medical image computing and computer-assisted intervention. pp. 690–700. Springer (2024)
13. Richardson, D.B., Cardis, E., Daniels, R.D., Gillies, M., O'Hagan, J.A., Hamra, G.B., Haylock, R., Laurier, D., Leuraud, K., Moissonnier, M., et al.: Risk of cancer from occupational exposure to ionising radiation: retrospective cohort study of workers in france, the united kingdom, and the united states (inworks). Bmj **351** (2015)
14. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
15. 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)
16. Tapp, A., Parida, A., Zhao, C., Lam, V., Lepore, N., Anwar, S.M., Linguraru, M.G.: Mr to ct synthesis using 3d latent diffusion. In: 2024 IEEE International Symposium on Biomedical Imaging (ISBI). pp. 1–5 (2024)
17. Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C.: N4itk: improved n3 bias correction. IEEE transactions on medical imaging **29**(6), 1310–1320 (2010)
18. Vajpayee, R., Agrawal, V., Krishnamurthi, G.: Structurally-constrained optical-flow-guided adversarial generation of synthetic ct for mr-only radiotherapy treatment planning. Scientific Reports **12**(1), 14855 (2022)
19. Wang, Z., Jiang, J.X., Wang, H.R., Zhou, L., Li, Y.: Localized adaptive style mixing for feature statistics manipulation in medical image translation with limited data. Expert Systems with Applications **277**, 127217 (2025)
20. Wolny, A., Cerrone, L., Vijayan, A., Tofanelli, R., Barro, A.V., Louveaux, M., Wenzl, C., Strauss, S., Wilson-Sánchez, D., Lymbouridou, R., et al.: Accurate and versatile 3d segmentation of plant tissues at cellular resolution. Elife **9**, e57613 (2020)

21. Yin, C., Hu, P., Qin, L., Wang, Z., Zhao, H.: The current status and future directions on nanoparticles for tumor molecular imaging. International Journal of Nanomedicine **19**, 9549–9574 (2024), pMID: 39296941
22. Zhong, L., Chen, Z., Shu, H., Zheng, K., Li, Y., Chen, W., Wu, Y., Ma, J., Feng, Q., Yang, W.: Multi-scale tokens-aware transformer network for multi-region and multi-sequence mr-to-ct synthesis in a single model. IEEE Transactions on Medical Imaging **43**(2), 794–806 (2024)