MRI-to-Synthetic CT for Intracranial Stereotactic Radiotherapy: a Comparative Study of Deep Learning Architectures with Focus on Bone Fidelity

Danilo Paglialunga, Alfonso Belardo, Maria Giulia Ubeira Gabellini*, Simone Raggio, Sara Broggi, Gabriele Palazzo, Bianca Bordigoni, Gianluca Gragnaniello, Antonella del Vecchio, Lorenzo Placidi and Claudio Fiorino

Medical Physics Unit, IRCCS Ospedale San Raffaele, Milan, Italy
* corresponding author, ubeira.mariagiulia@hsr.it

Abstract. Intracranial stereotactic radiation therapy (SRS/SRT) with robotic systems demands sub-millimetric geometric accuracy. As image guidance relies on Digitally Reconstructed Radiographs (DRRs) derived from the planning CT for skull tracking along the bone-tissue gradient, accurate reconstruction of this interface drives treatment precision. MRI-only workflows would remove the MRI-CT registration error, but they shift onto the synthesis model the full responsibility of reproducing correct spatial information and Hounsfield Unit (HU) values. We present a modular pipeline for MRI-to-synthetic-CT (sCT) generation, built around a 3D Attention U-Net trained on the SynthRAD2023 Brain dataset (180 brain MRI/CT pairs). The pipeline combines additive attention gates on skip connections with a bone-weighted composite loss, a cranial-specific HU normalization window ($[-1000, 2000]$, bone defined above 300 HU), bone-aware patch sampling, Exponential Moving Average (EMA) of the weights, and Hann-weighted sliding window inference. We compared four architectures under one shared protocol: a conditional GAN, a denoising diffusion probabilistic model (DDPM), a standard 3D U-Net, and the proposed 3D Attention U-Net. Attention U-Net provided the best HU accuracy, with a global MAE of ~ 106 HU and a bone-region MAE of ~ 195 HU. A per-patient reliability assessment module then combines bone geometry, DRR consistency, and HU accuracy into a single confidence score, providing a structured retrospective framework for assessing case-wise suitability in stereotactic workflows.

Keywords: Synthetic CT · MRI-only Radiotherapy · Stereotactic Radiotherapy · 3D Attention U-Net · CyberKnife · Deep Learning

1 Introduction

Intracranial stereotactic radiosurgery (SRS) and stereotactic radiotherapy (SRT) deliver ablative doses in one or few fractions, with steep dose gradients and margins often around or below a millimeter [4]. Two requirements follow: precise

geometric targeting and correct Hounsfield Unit (HU) values for dose calculation. Computed tomography (CT) gives direct HU values but poor soft tissue contrast, while Magnetic Resonance Imaging (MRI) does the opposite, delineating soft tissue and organs at risk better but carrying no direct HU information [1]. The standard SRS workflow acquires CT and MRI separately and registers them rigidly, introducing an inter-modality error that propagates into contouring and dose calculation and is costly in stereotaxis. MRI-only workflows avoid this [2]: a synthetic CT (sCT) generated directly from the MRI replaces the planning CT, removing the registration step and the extra dose from a separate CT scan. The trade-off is that the synthesis model now carries the burden of reconstructing accurate HU values, a task prone to model-dependent artifacts [3]. Brain sCT studies typically report a global Mean Absolute Error (MAE) between 50 and 90 HU in-distribution data, corresponding to dose differences generally below 2% relative to CT-based calculations [4, 5]. SynthRAD challenges provided shared datasets and protocols and showed that the largest residual errors are at the air and bone-tissue interfaces [8, 9]. Cortical bone remains the hardest part: MRI gives almost no signal over compact bone, with no natural zero reference as in HU, so models trained with generic losses tend to pull bone values toward the soft tissue range, underestimating bone HU and beam attenuation along paths crossing the skull [10, 11]. The DRRs from the planning CT serve as a geometric reference for robotic SRS/SRT delivered with the CyberKnife image-guidance system: DDRs are compared to in-room kV images acquired during treatment for bone tracking and 6-DoF positioning corrections [12]. Consequently, any error in bone representation or geometric consistency of CT-MRI feeds into positioning accuracy and dose delivery [13]. In the current work, we build and evaluate a modular pipeline for brain MRI-to-sCT synthesis based on a *3D Attention U-Net*, rather than proposing a new network family. Main aims of the study were to: (1) develop a pipeline with identical pre- and post-processing at training and inference, reducing domain-shift risk; (2) use a bone-weighted composite loss function; (3) introduce a multi-gate, per-patient reliability score that combines bone geometry, DRR consistency, and HU accuracy into one number; and (4) to carry out a head-to-head comparison with four alternative architectures in a common evaluation cohort.

2 Materials and Methods

2.1 Dataset

Models were trained and validated on the publicly available SynthRAD2023 Brain dataset [8], providing paired T1-weighted MRI/CT acquisitions (144 training, 36 validation). In addition, for independent evaluation, we used a fully anonymized institutional cohort of 10 SRS/SRT patients, imaged with contrast-enhanced T1-weighted (T1-GD) MRI and planning CT; this contrast protocol differs from the non-contrast T1w training data, a domain shift discussed in Section 4.

2.2 Pre-processing

The same procedures for volume pre-processing were followed at training and inference. MRI volumes were bias-corrected with N4 [30], resampled to 1.0 mm isotropic voxels, normalized z-score within a body mask and clipped to $[-5, 5]$ standard deviations. Volumes were cropped to the body bounding box with a margin of 15 mm. The reference CT was rigidly registered and resampled onto the MRI grid and mapped to a normalized target:

$$y_{\text{norm}} = \text{clip}\left(2\,\frac{\text{HU} - \text{HU}_{\text{min}}}{\text{HU}_{\text{max}} - \text{HU}_{\text{min}}} - 1,\ -1,\ 1\right), \tag{1}$$

with the inverse applied at inference. We use a brain-adapted window of $[-1000, 2000]$ HU, since cortical bone rarely exceeds $\sim$2000 HU; a threshold of 300 HU defines bone for the loss and geometry metrics below.

2.3 Network Architectures

We benchmark four competitive architectures, inspired by top-performing designs from the SynthRAD2023 Grand Challenge [8, 9]: a plain 3D residual U-Net, a conditional GAN, a DDPM cascade, and our proposed 3D Attention U-Net. All four networks map a single-channel MR patch to a single-channel sCT patch in $[-1, 1]$ through a final tanh. Our proposed *3D Attention U-Net* [18] is a three-level residual U-Net with additive attention gates on every skip connection (Fig. 1): the encoder characteristic $\mathbf{x}$ and the decoder gating signal $\mathbf{g}$ are combined into a per-voxel coefficient $\alpha = \sigma(\mathbf{W}_\psi\, \phi(\text{GN}(\mathbf{W}_x\mathbf{x} + \mathbf{W}_g\mathbf{g})))$, with ϕ the LeakyReLU and σ the sigmoid, and the gated skip $\alpha \odot \mathbf{x}$ replaces the plain skip. This steers the decoder toward the bone/soft-tissue interfaces and is where its gain over the plain U-Net comes from. Downsampling uses strided convolutions rather than pooling; group normalization [24] replaces batch normalization for a batch size of one; the bottleneck adds Dropout3D ($p = 0.1$) for regularization.

The *plain U-Net* shares this backbone without the attention gates, extending the 2D U-Net [16] to volumetric data following Çiçek et al. [17], isolating the contribution of gates. *cGAN* pairs a SwinUNETR generator [19] with a 3D PatchGAN discriminator under a least-squares (LSGAN) objective [20, 21], in line with the established adversarial MR-to-CT synthesis [22, 23]. *DDPM* is a two-stage cascade: our pre-trained Attention U-Net gives a coarse sCT, and a denoiser refines the residual under ϵ-prediction [25], with the cosine schedule of Nichol and Dhariwal [26] and DDIM sampling ($\eta = 0$, 50 steps) [27], following recent transformer-conditioned diffusion approaches to sCT synthesis [6, 28], but replacing the Swin-transformer backbone with our Attention U-Net as denoiser.

All four networks use bone-aware patch sampling (about half the patches guaranteed to contain bone) and MRI-only augmentation (flips, noise, intensity and gamma jitter). The U-Net variants and the cGAN generator were trained using AdamW [31] (learning rate 2×10^{-4}, cosine decay to 300 epochs), while the DDPM denoiser is trained on the frozen U-Net first-stage with a lower learning rate (2×10^{-5}) for 4 million optimization steps. An EMA of the weights [32] is kept throughout and used only for validation and inference.

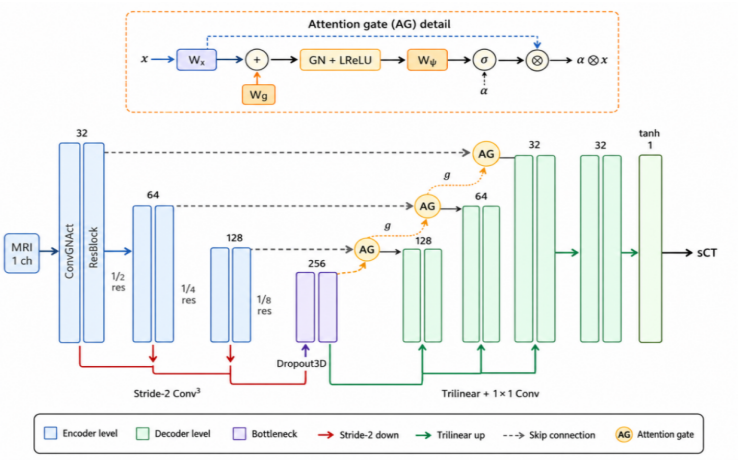

Fig. 1. 3D Attention U-Net: attention gates on each skip connection weight encoder features using the decoder gating, emphasizing the skull. Adapted from Oktay et al. [18].

2.4 Loss Functions

The plain and Attention U-Nets, and the pixel-wise term of the cGAN generator, share a composite reconstruction loss, computed in the normalized space of Eq. (1):

$$\mathcal{L}_{\text{reg}} = w_\ell \, \mathcal{L}_{\text{L1}} + w_b \, \mathcal{L}_b + w_s \, \mathcal{L}_{\text{SSIM}} + w_g \, \mathcal{L}_{\text{grad}} + w_c \, \mathcal{L}_{\text{region}}. \tag{2}$$

$\mathcal{L}_{\text{L1}}$, $\mathcal{L}_b$, $\mathcal{L}_{\text{SSIM}}$, and $\mathcal{L}_{\text{grad}}$ are, respectively: a global term ℓ_1; the same term restricted to bone voxels ($y_v > 300\,\text{HU}$) of the highest weight, since bone occupies little volume; a 3D structural term for soft-tissue contrast; and a gradient-matching term that sharpening anatomical boundaries. $\mathcal{L}_{\text{region}}$ penalizes the mean signed error within air, soft tissue, and bone separately, so a low global error cannot mask a class-wise bias.

The cGAN generator adds a least-squares adversarial term to $\mathcal{L}_{\text{reg}}$, with the discriminator trained under the matching LSGAN loss. The DDPM denoiser trains separately in the noise-prediction space, with a region-weighted ϵ-prediction loss that increases the weight of the bone and brain parenchyma over soft tissue and air.

2.5 Quality-Gated Inference and Reliability

Each volume was reconstructed at inference by Hann-weighted sliding-window aggregation ($96 \times 128 \times 128$ voxels, 50% overlap) within the body mask, then mapped back to HU via the inverse of Eq. (1). The sCT can be exported as a DICOM series for facilitating downstream integration into standard treatment-planning workflows.

Because image-quality metrics alone are only a moderate proxy for dosimetric and geometric validity [9], every sCT is checked against three gates. **Gate 1, bone geometry**: sCT and reference masks, both thresholded at 300 HU (Sect. 2.2), compared by Dice, ASSD, HD95; highest weight, since skull geometry drives dosimetric fidelity. **Gate 2, DRRs consistency**: DRRs from both volumes at four angles (two axes), compared by gradient-weighted normalized cross-correlation (gNCC), sensitive to the displacement CyberKnife tracking depends on. **Gate 3, HU accuracy**: global and per-tissue (air, soft tissue, bone) MAE, bias, RMSE, PSNR, SSIM, MS-SSIM [29] and Pearson correlation. The three scores were combined into a composite confidence $C = 100\,(w_B s_B + w_D s_D + w_H s_H)$, with $w_B = 0.45$, $w_D = 0.40$, $w_H = 0.15$, and $s_B, s_D, s_H \in [0, 1]$ per-gate scores clipped to that range. An sCT is declared reliable when $C \geq 80$ and no hard failure is triggered: a bone Dice below 0.75 or a mean gNCC below 0.65 renders the case unreliable regardless of C.

3 Results

All four models were evaluated on the independent institutional cohort ($n = 10$) according to the metrics in Section 2.5; median [IQR] metrics are reported in Table 1 while per-tissue distributions in Fig. 2. Composite confidence C and binary decision are the primary criteria, with image-quality metrics as support.

Table 1. Gate-level comparison on the independent test cohort ($n = 10$, median [IQR]). Best value per metric in **bold**.

	Gate 1, Bone geometry			Gate 2	Gate 3, HU accuracy				Reliab.
Method	Dice↑	ASSD [mm]↓	HD95 [mm]↓	gNCC↑	MAE [HU]↓	RMSE [HU]↓	SSIM↑	MS-SSIM↑	Pass
cGAN	0.78 [0.76,0.82]	1.05 [0.78,1.16]	3.67 [2.91,4.18]	0.79 [0.73,0.82]	126.1 [118.5,139.1]	252.0 [239.6,280.9]	0.68 [0.64,0.71]	0.90 [0.88,0.92]	1/10
DDPM	0.81 [0.78,0.83]	0.92 [0.81,1.07]	3.30 [2.59,3.71]	0.81 [0.78,0.84]	171.8 [165.0,182.7]	263.1 [250.8,282.4]	0.61 [0.57,0.65]	0.89 [0.89,0.91]	0/10
UNet3D	0.87 [0.84,0.89]	0.69 [0.48,0.83]	2.34 [1.73,3.19]	0.86 [0.82,0.89]	107.9 [97.0,117.3]	228.1 [197.8,257.9]	0.70 [0.67,0.74]	0.91 [0.90,0.93]	6/10
UNet3DAttn	**0.87** [0.84,0.89]	**0.64** [0.48,0.76]	**2.34** [1.80,2.96]	**0.88** [0.85,0.89]	**105.8** [91.8,116.9]	**217.4** [191.0,257.0]	**0.72** [0.71,0.76]	**0.93** [0.92,0.96]	**8/10**

3.1 Overall Performance Across Architectures

Across all three gates, the 3D Attention U-Net gives the most consistent profile, leading every metric, except HD95 and Dice, tied with the plain U-Net. The improvement over the plain U-Net in HU accuracy is narrow, confirming that the residual backbone alone is a strong regressor. The gap was larger in bone geometry and projection consistency. RMSE grows more than MAE for the plain U-Net, indicating localized errors in specific cases rather than a uniform HU shift. The lower certification rate (6/10 vs 8/10) is driven by cases shared across architectures: both non-certified Attention U-Net cases are among the plain U-Net's four, and all four fail Gate 1 (bone geometry), three outright. This suggests patient-specific anatomical difficulty in bone reconstruction rather than an attention-specific weakness. The cGAN and DDPM failed to certify for

different reasons. The cGAN reaches moderate MAE and RMSE (126.1 and 252.0 HU) but showed the weakest bone geometry and projection consistency. Only 1/10 cases are certified, since its adversarial objective yields plausible texture without the consistency that Gate 2 requires. The DDPM cascade is more instructive: its bone geometry and projection consistency are comparable to the regression models, well above the hard-failure floors, yet no case is certified (median $C = 53.7\%$). The reason is Gate 3: the DDPM shows by far the largest MAE and RMSE (171.8 and 263.1 HU). Its denoiser refines the coarse prediction, not the MRI, inheriting the training/test protocol gap noted in Section 2.1. The result is good structure but poor HU accuracy, exactly what a single image-quality score would miss and what motivates our explicit HU accuracy gate and hard-failure floors.

3.2 Tissue-Stratified and Reliability Analysis

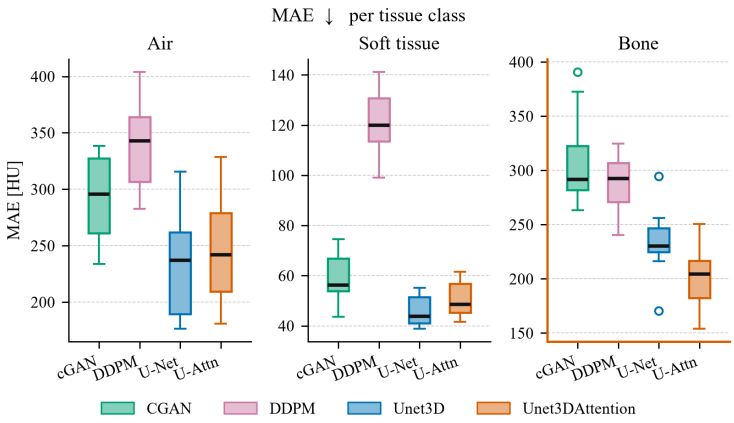

Fig. 2. Per-tissue MAE for Gate 3 (HU accuracy): the Attention U-Net gives the most compact bone-error distribution, the DDPM the largest variability.

Splitting Gate 3's HU accuracy according to the tissue class resulted in a consistent ranking (Fig. 2). The DDPM exhibited the highest and most dispersed errors in air and soft tissue and performed comparably to the cGAN in bone. Across all tissue classes, the cGAN showed errors larger than those of both U-Net variants. Although plain and Attention U-Nets perform similarly in air and soft tissue, Attention U-Net achieves a narrower error distribution in bone, reflecting the intended effect of its attention mechanism. These findings align with recent studies showing that regression-based networks with explicit bone supervision generally outperform adversarial and diffusion-based approaches [14, 15]. Bone remained as the tissue to be mimic for each architecture: even the Attention U-Net's bone MAE, the lowest of the four, remains above its own soft-tissue and

air errors. This reflects the well-documented tendency to regress toward the soft tissue range where compact bone gives little signal [10, 11]; our bone-stratified loss attenuates, but does not remove this gap. Looking to individual patient data, the plain U-Net's non-certified cases are driven mainly by Gates 1 and 2, consistent with Gate 3's lower weight.

3.3 Qualitative Evaluation

Complementing the cohort-level statistics, Fig. 3 compares the four architectures on patient 6, where all four perform best. The Attention U-Net gives the sharpest, most complete skull outline, closely tracking the reference contour at the inner and outer tables. The plain U-Net is close behind, with a slightly thicker, less continuous contour. cGAN and DDPM both blur the skull: fine bone structure is smoothed away and the contour drifts from the reference, matching their higher bone MAE in Fig. 2.

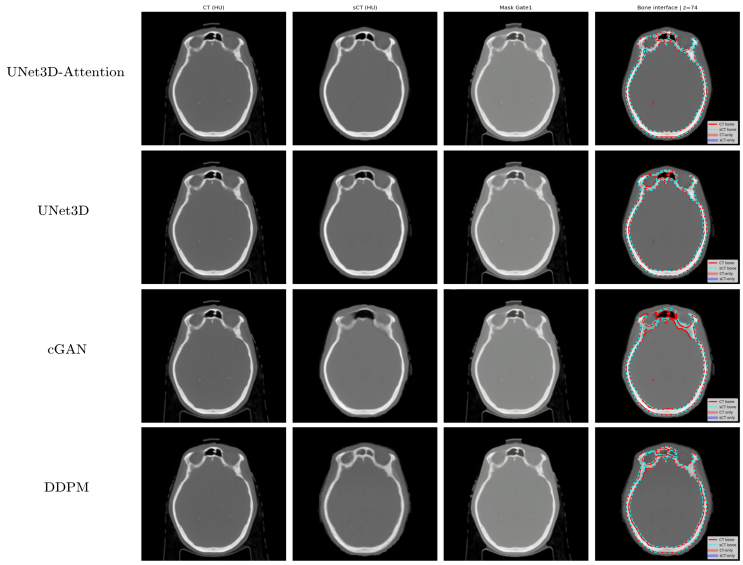

Fig. 3. Qualitative comparison for patient 6, where all four architectures perform best. Per architecture: mid-axial CT slice, sCT slice, bone mask and bone-contour overlay.

4 Discussion

A 3D Attention U-Net with bone-oriented composite loss and an adapted HU window is a robust choice for brain MRI-to-sCT synthesis, outperforming the plain U-Net, cGAN, and diffusion cascade on error and structural metrics, mainly

via reduced RMSE in bone/soft-tissue regions and a moderate median-MAE shift. Our MAE in-distribution validation (∼55 HU) falls within the ∼50–90 HU range for brain sCT [4, 5, 8, 9], while the higher held-out error (∼106 HU) reflects the protocol shift (T1-GD vs T1w). Brain and SRS-specific sCT work is moving quickly: transformer-conditioned diffusion models report HU accuracy close to ours [6], and single-MRI residual U-Nets have been proposed for frameless radiosurgery [7]. The novelty of our methodology stays in the training tuned onto cortical bone: the bone-weighted loss, the cranial HU window and bone-aware sampling target the tissue CyberKnife's DRR tracking depends on, consistent with external comparisons where regression networks with explicit supervision beat adversarial and diffusion alternatives [14, 15]. The SynthRAD benchmark keeps evolving: a 2025 release covers the head-and-neck, thorax, and abdomen [33], not the brain, so SynthRAD2023 remains our reference. The results of this challenge show why bone-first design matters for CyberKnife stereotaxis: the DDPM matches the regression models on bone geometry and projection consistency, but it is uncertified because of its large HU error. A pipeline relying on one perceptual metric or error statistic would misrank it. Brain anatomy is mostly soft tissue and bone with few air cavities, so bounded HU errors cause small dose differences along CyberKnife's beam paths, while geometric errors affect DRR quality, bone tracking and consequently the accuracy of the delivery. Weighting bone geometry and DRR consistency above global HU statistics reflects the accuracy requirements of SRS/SRT rather than modeling convenience. Our study has several limitations. Clinically relevant dosimetric validity [2] remains to be assessed in future work. The DDPM cascade depends on the Attention U-Net output, inheriting errors and limiting generalization. The institutional test cohort is limited to ten patients, warranting caution when interpreting HU differences, although improvement in bone-related performance remains consistent. No site-specific fine-tuning was performed, so held-out errors are a lower bound on achievable performance. Cortical bone remains the hardest tissue for every architecture. Near-zero MRI signal over compact bone is a structural limit that loss stratification attenuates but does not remove.

5 Conclusions

We presented a modular pipeline for brain MRI-to-sCT generation in intracranial SRS/SRT, built around a 3D Attention U-Net with a bone-oriented composite loss, an adapted HU window, and a per-patient reliability module. On an independent cohort of ten SRS/SRT patients, it achieves the best HU accuracy of the four architectures (global MAE ∼106 HU, bone MAE ∼195 HU) and certifies 8/10 cases as reliable. Dosimetric validation on this cohort, together with its planned expansion to ∼80 patients, is the next step toward clinical deployment.

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

Funding. The study was supported by an AIRC (Associazione Italiana per la Ricerca sul Cancro) grant (IG 2025 ID32254).

References

1. Edmund, J.M., Nyholm, T.: A review of substitute CT generation for MRI-only radiation therapy. Radiat. Oncol. **12**, 28 (2017)
2. Villegas, F., Dal Bello, R., Alvarez-Andres, E., Dhont, J., Janssen, T., Milan, L., Robert, C., Salagean, G.A.M., Tejedor, N., Trnková, P., Fusella, M., Placidi, L., Cusumano, D.: Challenges and opportunities in the development and clinical implementation of artificial intelligence based synthetic computed tomography for magnetic resonance only radiotherapy. Radiother. Oncol. **198**, 110387 (2024)
3. Owrangi, A.M., Greer, P.B., Glide-Hurst, C.K.: MRI-only treatment planning: benefits and challenges. Phys. Med. Biol. **63**(5), 05TR01 (2018)
4. Kazemifar, S., McGuire, S., Timmerman, R., Wardak, Z., Nguyen, D., Park, Y., Jiang, S., Owrangi, A.: MRI-only brain radiotherapy: assessing the dosimetric accuracy of synthetic CT images generated using a deep learning approach. Radiother. Oncol. **136**, 56-63 (2019)
5. Dinkla, A.M., Wolterink, J.M., Maspero, M., et al.: MR-only brain radiation therapy: dosimetric evaluation of synthetic CTs generated by a dilated convolutional neural network. Int. J. Radiat. Oncol. Biol. Phys. **102**(4), 801-812 (2018)
6. Pan, S., Abouei, E., Wynne, J., Chang, C.W., Wang, T., Qiu, R.L.J., Li, Y., Peng, J., Roper, J., Patel, P., Yu, D.S., Mao, H., Yang, X.: Synthetic CT generation from MRI using 3D transformer-based denoising diffusion model. Med. Phys. **51**(4), 2538-2548 (2024)
7. Shu, X., Lu, K., Zhao, J., Ginn, J., Kim, Y., Yang, Z., Adamson, J., Mullikin, T., Wang, C.: Knowledge-based deep residual U-Net (DRU) for synthetic CT generation using a single MR volume for frameless radiosurgery. J. Appl. Clin. Med. Phys., early view (2025). https://doi.org/10.1002/acm2.70343
8. Thummerer, A., van der Bijl, E., Galapon, A. Jr., Verhoeff, J.J.C., Langendijk, J.A., Both, S., van den Berg, C.A.T., Maspero, M.: SynthRAD2023 Grand Challenge dataset: generating synthetic CT for radiotherapy. Med. Phys. **50**(7), 4664-4674 (2023)
9. Huijben, E.M.C., Terpstra, M.L., Pai, S., Thummerer, A., et al.: Generating synthetic computed tomography for radiotherapy: SynthRAD2023 challenge report. Med. Image Anal. **97**, 103276 (2024)
10. Wang, T., Lei, Y., Fu, Y., Wynne, J.F., Curran, W.J., Liu, T., Yang, X.: Medical imaging synthesis using deep learning and its clinical applications: a review. J. Appl. Clin. Med. Phys. **22**(1), 11-36 (2021)
11. Bambach, S., Ho, M.L.: Deep learning for synthetic CT from bone MRI in the head and neck. AJNR Am. J. Neuroradiol. **43**(8), 1172-1179 (2022)
12. Fu, D., Kuduvalli, G.: A fast, accurate, and automatic 2D-3D image registration for image-guided cranial radiosurgery. Med. Phys. **35**(5), 2180-2194 (2008)
13. Jiao, S., Zhao, X., Zhou, P., Geng, M.: Technical note: MR image-based synthesis CT for CyberKnife robotic stereotactic radiosurgery. Biomed. Phys. Eng. Express **10**, 057001 (2024)
14. Honey, E., Helbo, A., Petersen, J.: Comparative analysis of GAN and diffusion for MRI-to-CT translation. arXiv:2509.22049 (2025)
15. Gholamiankhah, F., Mostafapour, S., Arabi, H.: Deep learning-based synthetic CT generation from MR images: comparison of generative adversarial and residual neural networks. Int. J. Radiat. Res. **20**(1), 121-130 (2022)
16. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: MICCAI 2015. LNCS, vol. 9351, pp. 234-241. Springer (2015)

17. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: MICCAI 2016. LNCS, vol. 9901, pp. 424-432. Springer (2016)
18. Oktay, O., Schlemper, J., Le Folgoc, L., et al.: Attention U-Net: learning where to look for the pancreas. In: MIDL (2018). arXiv:1804.03999
19. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H.R., Xu, D.: Swin UNETR: Swin transformers for semantic segmentation of brain tumors in MRI images. In: BrainLes, MICCAI 2021 Workshop. LNCS, vol. 12962, pp. 272-284. Springer (2022)
20. Mao, X., Li, Q., Xie, H., Lau, R.Y.K., Wang, Z., Smolley, S.P.: Least squares generative adversarial networks. In: ICCV, pp. 2794-2802 (2017)
21. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: CVPR, pp. 5967-5976 (2017)
22. Emami, H., Dong, M., Nejad-Davarani, S.P., Glide-Hurst, C.K.: Generating synthetic CTs from magnetic resonance images using generative adversarial networks. Med. Phys. **45**(8), 3627-3636 (2018)
23. Maspero, M., Savenije, M.H.F., Dinkla, A.M., et al.: Dose evaluation of fast synthetic-CT generation using a generative adversarial network for general pelvis MR-only radiotherapy. Phys. Med. Biol. **63**(18), 185001 (2018)
24. Wu, Y., He, K.: Group normalization. In: ECCV. LNCS, vol. 11217, pp. 3-19. Springer (2018)
25. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. In: NeurIPS, vol. 33, pp. 6840-6851 (2020)
26. Nichol, A.Q., Dhariwal, P.: Improved denoising diffusion probabilistic models. In: ICML, pp. 8162-8171 (2021)
27. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. In: ICLR (2021)
28. Moradi, N., et al.: VS-DDPM: efficient low-cost diffusion model for medical modality translation. arXiv:2604.22942 (2026)
29. Wang, Z., Simoncelli, E.P., Bovik, A.C.: Multiscale structural similarity for image quality assessment. In: 37th Asilomar Conf. Signals, Systems and Computers, pp. 1398-1402 (2003)
30. Tustison, N.J., Avants, B.B., Cook, P.A., et al.: N4ITK: improved N3 bias correction. IEEE Trans. Med. Imaging **29**(6), 1310-1320 (2010)
31. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. In: ICLR (2019)
32. Polyak, B.T., Juditsky, A.B.: Acceleration of stochastic approximation by averaging. SIAM J. Control Optim. **30**(4), 838-855 (1992)
33. Thummerer, A., van der Bijl, E., Galapon, A. Jr., Kamp, F., Savenije, M., et al.: SynthRAD2025 Grand Challenge dataset: generating synthetic CTs for radiotherapy from head to abdomen. Med. Phys. **52**(7), e17981 (2025)