

Article

Not peer-reviewed version

Contrast-enhanced Liver MR Synthesis using Gradient Regularized Multi-Modal Multi-Discrimination Sparse Attention Fusion GAN

[Changzhe Jiao](#) , Diane Ling , Shelly Bian , April Vasantachart , Karen Cheng , Shahil Mehta , Derrick Lock , [Zhenyu Zhu](#) , Mary Feng , Horatio Thomas , Jessica E Scholey , [Ke Sheng](#) , Zhaoyang Fan , [Wensha Yang](#) *

Posted Date: 10 May 2023

doi: 10.20944/preprints202305.0705.v1

Keywords: MR synthesis, GAN, multi-modal fusion, tumor monitoring, contrast enhancement



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Contrast-Enhanced Liver MR Synthesis Using Gradient Regularized Multi-Modal Multi-Discrimination Sparse Attention Fusion GAN

Changzhe Jiao ^{1,2}, Diane Ling ¹, Shelly Bian ¹, April Vasantachart ¹, Karen Cheng ¹, Shahil Mehta ¹, Derick Lock ¹, Zhenyu Zhu ³, Mary Feng ², Horatio Thomas ², Jessica E. Scholey ², Ke Sheng ², Zhaoyang Fan ⁴ and Wensha Yang ^{1,2,*}

¹ Department of Radiation Oncology, Keck School of Medicine of USC, Los Angeles, CA 90033

² Department of Radiation Oncology, UC San Francisco, San Francisco, CA 94143

³ Guangzhou Institute of Technology, Xidian University, Guangzhou, China

⁴ Department of Radiology, Keck School of Medicine of USC, Los Angeles, CA 90033

* Correspondence: Wensha.Yang@ucsf.edu

Abstract: Purposes: To provide abdominal contrast-enhanced MR image synthesis, we developed an image gradient regularized multi-modal multi-discrimination sparse-attention fusion generative adversarial network (GRMM-GAN) to avoid repeated contrast injections to patients and facilitate adaptive monitoring. Methods: With IRB approval, 165 abdominal MR studies from 61 liver cancer patients were retrospectively solicited from our institutional database. Each study included T2, T1 pre-contrast (T1pre), and T1 contrast-enhanced (T1ce) images. The GRMM-GAN synthesis pipeline consists of a sparse attention fusion network, an image gradient regularizer (GR), and a generative adversarial network with multi-discrimination. The studies were randomly divided into 115 for training, 20 for validation, and 30 for testing. The two pre-contrast MR modalities, T2 and T1pre images, were adopted as inputs in the training phase. The T1ce image at the portal venous phase was used as an output. The synthesized T1ce images were compared with the ground truth T1ce images. The evaluation metrics include peak signal-to-noise ratio (PSNR), structural-similarity-index (SSIM), and mean-squared-error (MSE). A Turing test and experts' contours evaluated the image synthesis quality. Results: The proposed GRMM-GAN model achieved a PSNR of 28.56, an SSIM of 0.869, and an MSE of 83.27. The proposed model showed statistically significant improvements in all metrics tested with p-values <0.05 over the state-of-the-art model comparisons. The average Turing test score was 52.33%, close to random guessing, supporting the model's effectiveness for clinical application. In the tumor-specific region analysis, the average tumor contrast-to-noise ratio (CNR) of the synthesized MR images was not statistically significant from the real MR images. The average DICE from real vs. synthetic images was 0.90, compared to the inter-operator DICE of 0.91. Conclusion: We demonstrated the function of a novel multi-modal MR image synthesis neural network, GRMM-GAN, for T1ce MR synthesis based on pre-contrast T1 and T2 MR images. GRMM-GAN shows promise for avoiding repeated contrast injections during radiation therapy treatment.

Keywords: MR synthesis; GAN; multi-modal fusion; tumor monitoring; contrast enhancement

1. Introduction

The American Cancer Society has estimated 41,260 newly diagnosed liver cancer and intrahepatic bile duct cancer patients in the United States and 30,520 related deaths [1]. Medical imaging is essential in both the diagnosis and treatment of liver cancer. Magnetic resonance (MR) provides excellent soft tissue contrast with its versatile and functional imaging sequences. Compared to computed tomography (CT), multi-phase contrast-enhanced MR has shown improved sensitivity

in detecting hepatocellular carcinoma (HCC), the most common type of primary liver cancer [2]. Hence, several guidelines recommend multi-phase contrast-enhanced MR as the standard imaging modality for liver cancer diagnosis [3–6]. Treatments of liver cancer include surgery, liver transplant, thermal ablation, chemo-radioembolization, external beam radiation therapy (EBRT), targeted drug therapy, or immunotherapy. Within these modalities, EBRT has the advantage of being geometrically targeted and non-invasive. However, the outcome of conventional fractionated EBRT is unsatisfactory due to tumor radioresistance and the risk of radiation-induced liver disease [7]. Stereotactic body radiation therapy (SBRT), a more precise, ablative type of radiation, has gained popularity in the recent decade in overcoming the radioresistance of various cancers, including liver cancer [8]. SBRT is an aggressive form of EBRT delivering highly hypo fractionated, thus more biologically potent doses. The success of SBRT hinges on geometrically accurate tumor targeting and rapid dose drop-off to spare the surrounding normal tissues. In other words, SBRT requires a more stringent tumor definition and reduced geometrical margin. To this end, multi-phase contrast-enhanced MR has been increasingly used to register with CT for SBRT planning to define tumor and normal anatomy interfaces better. However, deformable image registration between CT and MR can be error-prone [9].

More recently, MR-guided linear accelerators (LINACs) have been commercialized, making MR-only SBRT planning and adaptation an appealing alternative circumventing MR-CT registration. Besides being more sensitive in detecting liver tumors for more accurate treatment, contrast-enhanced MR imaging could be used for daily treatment response assessment by providing better visibility and richer details in the region of interest as the contrast agents flow with blood vessels over time. However, obtaining daily contrast-enhanced images on MR-LINAC can be prohibitive due to the increased risk of side effects from repeated contrast injections [10–12]. The side effects can be severe for patients with compromised kidney function, a condition more commonly observed in patients with liver cancer [13]. Safety concerns regarding administering gadolinium-based contrast agents, nephrogenic systemic fibrosis, and additional procedure time [14–16] also preclude more frequent contrast-enhanced MR acquisition. The limitation on contrast usage thus severely diminishes the value of contrast-enhanced MR in MR-guided radiotherapy for daily tumor targeting and early response assessment. Therefore, there is a clinical need for predicting contrast-enhancement MR information without repeated contrast injection.

Medical image synthesis [17] is a rapidly developing area benefitting from deep learning (DL) methods, among which the generative adversarial network (GAN) [18,19] was mainly designed for image synthesis. The state-of-the-art GAN methods include pix2pix [20], which enforced the L1 norm paired image similarity and was efficient in paired image synthesis, and CycleGan [21], which learned the high-level features in the source domain and then applied them for style translation. GAN medical imaging contributions include medical image enhancement [22–24], super-resolution [25,26], cross-modal MR-CT image synthesis [27,28], and multi-contrast MR synthesis [29–31].

Multi-modal learning has attracted increasing interest in MRI synthesis due to the available salient and mutual-complementary information. However, inherent redundancy and noisy artifacts exist across multiple modalities, making the efficacious fused learning from multi-modal MR images difficult. Several studies attempted to tackle the challenge. MM-Syns [32] learned a shared latent feature space for multi-modal data and synthesized the multi-output brain MR image via an encoder-decoder structure. Similarly, LR-cGAN [33] adopted the encoder-decoder design in a GAN-based model for brain MR image synthesis. Hi-Net [34] proposed a mixed fusion block with element-wise operations for multi-modal feature fusion and brain MR image synthesis. MMgSN-Net [35] adopted the structure of the Hi-Net with improved self-attention fusion for nasopharyngeal carcinoma MR image synthesis. MustGAN [36] designed a flexible multi-stream fusion framework to aggregate information across multiple sources and predicted missing modality data.

However, abdominal contrast-enhanced MR image synthesis from multi-modal inputs (e.g., T2, T1pre) is still challenging. To the best of our knowledge, there is currently no reported work for multi-modal contrast-enhanced liver MR image synthesis. We hypothesize that three difficulties account for this gap: 1) the inherent heterogeneity across different tumor types and patients in abdominal MR

images. Compared with MR image synthesis for other anatomical sites, there is more significant variation in abdominal MR image characteristics due to different tissue composition, vascularization, perfusion, and motion artifacts. The interpatient heterogeneity poses substantial challenges to the discriminative feature learning of the multi-modal fusion methods. 2) co-registration error across abdominal MR modalities. Inevitable breathing motion during the collection of abdominal MR images results in mismatches between different sequences. The uncertainties in the co-registration could confuse a synthesis method, especially for the deep learning model, because of strong memory. 3) blurring and over-smoothing effect of MR image synthesis. The blurring and over-smoothing effects of synthesized images are common and recognized as an issue in GAN-based models [37,38].

This study presents a novel image gradient-regularized multi-modal multi-discrimination sparse-attention fusion generative adversarial network (GRMM-GAN), a non-intrusive, efficient, and cost-saving clinical tool for contrast-enhanced abdominal MR image synthesis. The GRMM-GAN produces synthetic contrast-enhanced abdominal MR images to enable more accurate tumor delineation and response assessment for adaptive liver radiation therapy.

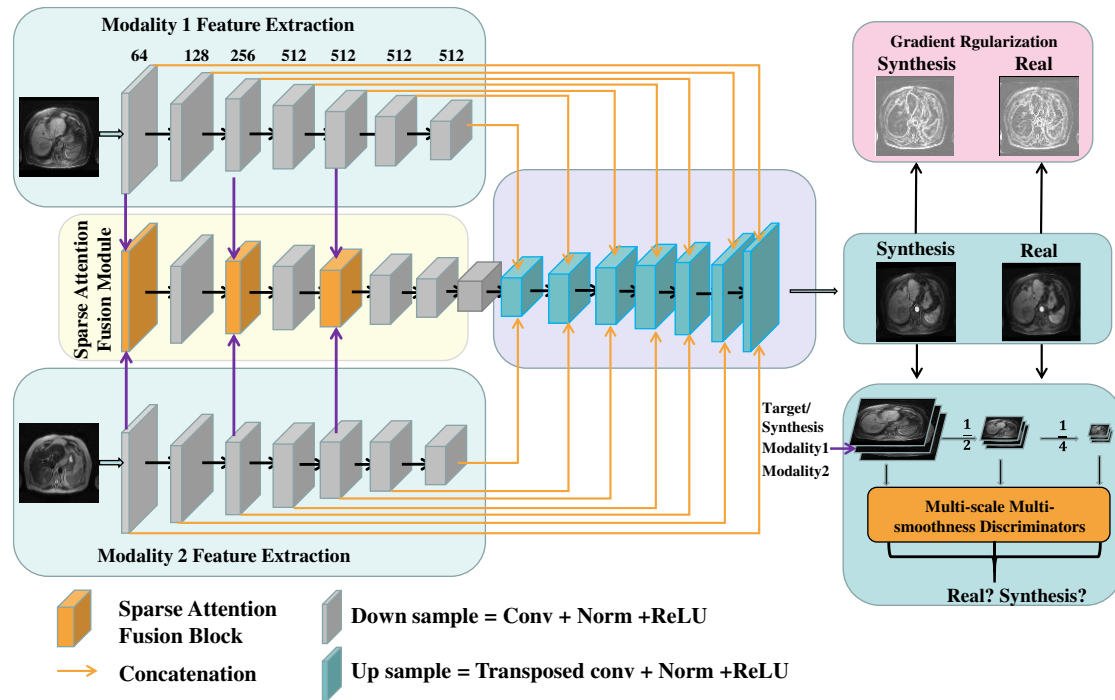
2. Methods

2.1. Data and preprocessing

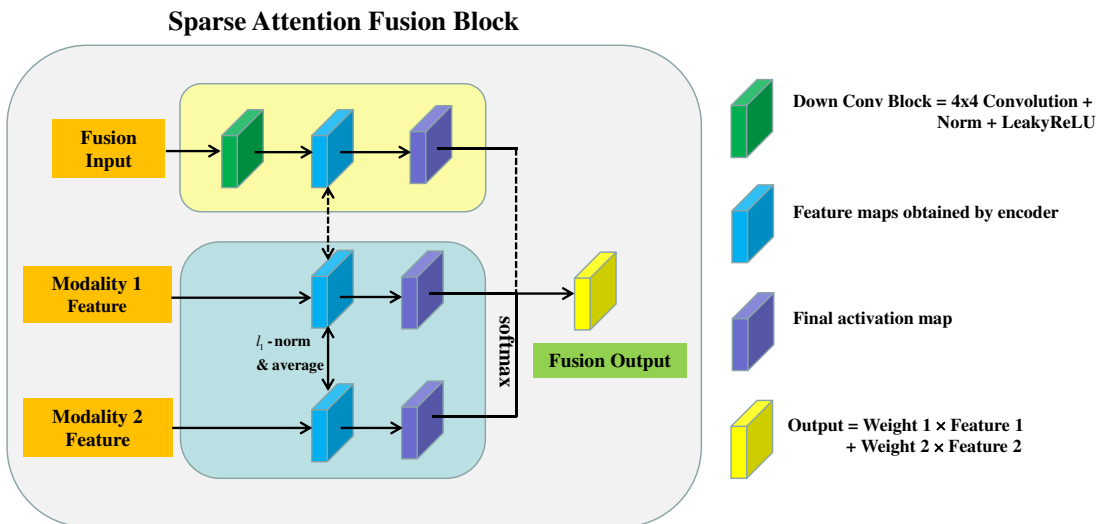
With IRB approval, we randomly solicited 165 MR studies of 61 liver patients from our institutional database. The patients' demographic and clinical informations are summarized in the Supplement Table S1. Each study included three modalities, T2, T1pre-contrast (T1pre), and T1 contrast-enhanced (T1ce) at the portal venous phase. The three modalities of each study were rigidly registered in VelocityAI™ (Varian, a Siemens Healthiness company, Palo Alto, CA). Various resolutions, including 640x640, 512x512, and 320x320, were used in the original MR images. In practice, we downsampled the MR images to 256x256 to balance computational cost and image quality.

2.2. Overall pipeline

The GRMM-GAN MR synthesis pipeline consists of a conditional GAN baseline and three modules: sparse attention fusion, gradient regularization (GR) mechanism, and multi-scale multi-smoothness discriminators (MMD), as shown in Figure 1 (a). Specifically, the sparse attention fusion module was designed to extract modal-specific features from each input modality (T1pre and T2) while eliminating the inherent redundancy and noisy artifacts and discovering the salient and mutual-complementary information. The image gradient regularization mechanism preserves the crucial texture and edge features of the abdominal organs and blood vessels such that the inherent heterogeneous knowledge of the abdominal MRs can be well studied. The multi-scale multi-smoothness discriminators examine the synthesis images in different scales and smoothness levels to overcome the blurring and over-smoothing effect. The following sections explain the baseline model and each network module in detail.



(a) Network design and structure



(b) Sparse fusion module

Figure 1. Framework of the proposed GR-MSSF GAN. (a) Network design and structure. (b) Sparse fusion module.

2.3. Baseline Conditional GAN model

The GRMM-GAN method is derived from a conditional GAN framework consisting of a generator G and a discriminator D . The generator's objective G is to fuse the multi-modal pre-contrast MR inputs, T1pre and T2, and predict the contrast-enhanced T1ce. At the same time, the

discriminator D focuses on the discrimination of the synthetic MR from the real MR. The prime conditional GAN objective function L_{cGAN} is given by

$$L_{cGAN}(G, D) = E_{x_1, x_2, y} [\log D(x_1, x_2, y)] + E_{x_1, x_2} \left[\log \left(1 - D(x_1, x_2, G(x_1, x_2)) \right) \right] \quad (1)$$

where the multi-modal inputs and target reference are denoted as x_1, x_2 and y , respectively.

Following the pix2pix method [20], the L1 loss is employed to promote the structural consistency between the real and the synthetic output and to avoid blurring. The L1 loss is defined as:

$$L_{L1}(G) = E_{x_1, x_2, y} [\|y - G(x_1, x_2)\|_1] \quad (2)$$

2.4. Sparse attention fusion

Although the current conditional GAN-based method shows promise for MR image synthesis, the inherent redundancy and noisy artifacts across multiple modalities present challenges for the effective fused learning of multi-modal abdominal MR images. To achieve superior fusion by discovering the salient and mutual-complementary information in multiple MR modalities while discriminating redundancy and noise from distortion and ambiguous co-registration, we designed a sparse attention fusion network, as shown in Figure 1 (b). This module was inspired by the previously proposed L₁ sparsity-regularized attention feature fusion work (L1-attention Fusion) [39]. Specifically, a sparse regularization term $\lambda \|c_i\|$ was introduced to the learned attention (c_i), where the sparsity regularizer (λ) controls the sparsity level of the attention weights. The sparse attention updates were realized using soft thresholding after the forward inference at each iteration, shown as follows:

$$c^* = S_\lambda(c), \quad \text{s.t. } S_\lambda(c) = \max(c - \lambda, 0) \quad (3)$$

The sparsity constraint, applied to the attention estimated for multi-modal fusion, eliminates the inherent redundancy across modalities and improves discriminative ability.

2.5. Gradient regularization mechanism

We added an image gradient regularization to manage the inter-patient heterogeneity and capture salient high-level features. At the same time, the network discriminated the redundancy and artifacts from the noisy inputs. The gradient regularization mechanism [40] introduced a gradient preservation loss shown as the following:

$$L_{GR}(G) = \|g \otimes G(x_1, x_2) - g \otimes y\|_2^2 = \|g \otimes (G(x_1, x_2) - y)\|_2^2 \quad (4)$$

where g is the Sobel gradient operator [41] and \otimes denotes the convolution operation, as shown in Figure 2 (a). A similar idea was recently applied as Ea-GANs for brain MR image synthesis [42]. The gradient regularization mechanism enforced the fidelity of the first-order information between the synthetic and the real T1ce images. It thus preserved the crucial texture and edge features of the abdominal organs and blood vessels while increasing the module robustness to heterogeneous redundancy and noise. Figure 2 (a) shows resultant gradient maps for the synthetic and the real T1ce images after the Sobel operation.

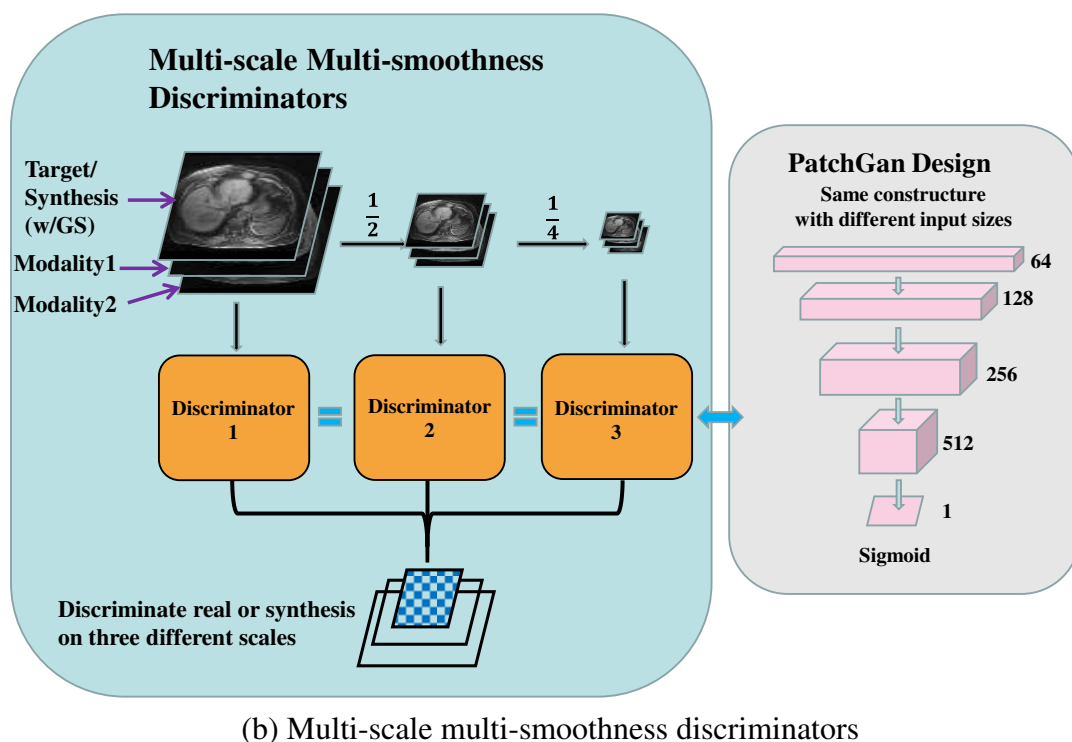
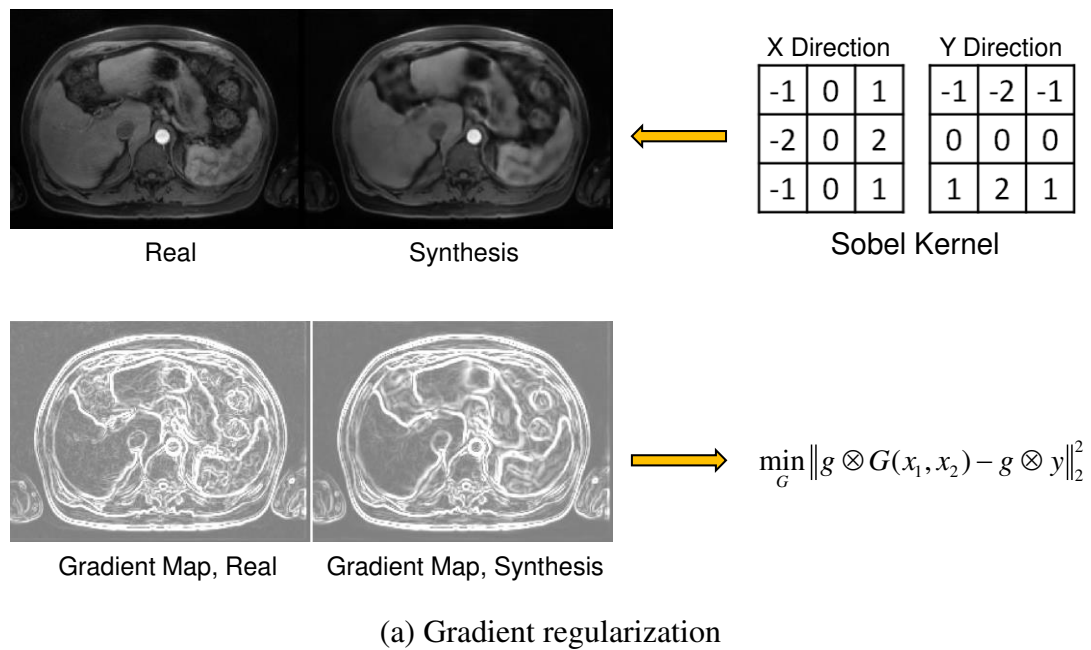


Figure 2. Illustration for the proposed gradient regularization and mutli-scale multi-smoothness discrimination mechanism. (a) Gradient regularization (GR). (b) Mutli-scale multi-smoothness discriminators (MMD).

2.6. multi-discrimination mechanism

In GAN-based models for image synthesis, blurring and over-smoothing effects are commonly observed since the optimization of the fidelity loss could easily fall on a local optimum [37,38]. This problem becomes more severe and visible due to the heterogeneous nature and ambiguous co-registration of multi-modal abdominal MR images. Inspired by the multi-scale discriminators [43,44] that enhanced the resolution of the image synthesis, as well as the illumination and scale-invariant SIFT feature learning [45,46], we propose multi-scale multi-smoothness discriminators (MMD) to

improve the discriminative ability and counter over-smoothness in synthesized images. The multi-modal inputs, synthesis, and ground truth images were downsampled twice into half of their original size. Then a Gaussian filter was applied to blur the synthetic image to simulate the over-smoothness effect. In total, there were three scales (256'256, 128'128, and 64'64), creating two pairs each with the original image downsampled over smoothed image for the discriminator. Three discriminators were trained given the paired blurred synthetic and real images, as shown in Figure 2 (b). The module thus learned blurring and over-smoothing effects in the synthesis and enforced the discrimination against these effects in adversarial training. The multi-scale multi-smoothness discrimination function is given as follows:

$$\min_G \max_{D_1, D_2, D_3} \sum_{k=1,2,3} L_{CGAN}(G, D_k) \quad (5)$$

where G is the generator for image synthesis, $D_k, k = 1,2,3$ are the three discriminators for over-smoothing discrimination, and L_{GAN} is the conditional GAN loss function.

2.7. Objective and Optimization

After incorporating the sparse attention fusion, image gradient regularization, and the multi-scale multi-smoothness discriminators into the prime conditional GAN model, we wrote the final objective of the proposed GRMM-GAN formula as follows:

$$\min_G \left(\left(\max_{D_1, D_2, D_3} \sum_{k=1,2,3} L_{CGAN}(G, D_k) \right) + \lambda_1 L_{GR}(G) + \lambda_2 L_{L_1}(G) \right) \quad (6)$$

where λ_1 and λ_2 are the trade-off hyper-parameters to balance the influence of different loss terms.

The proposed network was implemented on a workstation with Intel i9 7900x CPU and NVIDIA RTX 2080Ti'4 GPU, under the PyTorch 1.4 and Ubuntu 18.04 environment. For the contrast-enhanced MR image synthesis task, the training procedure of our proposed model takes 150 epochs, and we applied the Adam optimization algorithm with a batch size of 16 to update the network parameters. The balance hyper-parameter λ_2 was set to 100. λ_1 took a linear increase from 0 to 50 in the first 50 epochs and was frozen to 50 afterward. The learning rate for the first 50 epochs was fixed to 0.0002 and then linearly decreased to 0 in the following 100 epochs.

3. Model Evaluation

The performance of the proposed GRMM-GAN was evaluated quantitatively and compared with four state-of-the-art image synthesis networks: pix2pix [20], LR-cGAN [33], Hi-Net [34], and MMgSN-Net [35]. The pix2pix is the widely applied single modal image to image translation model, in which both T2 and T1pre were set as input. The other three comparisons are all state-of-the-art multi-modal MR image synthesis methods. In addition, six radiation oncologists (RadOnc 1-6), including two board-certified and four residents, performed Turing [47] tests to determine the authenticity of the synthesized T1ce images against the real images on 100 randomly (1:1 ratio) selected axial slices.

Three widely applied statistics metrics for evaluating the medical imaging synthesis include peak signal-to-noise ratio (PSNR), structural-similarity-index (SSIM), and mean-squared-error (MSE). The definitions of these metrics are presented as follows:

$$MSE = \frac{1}{N} (y - G(x))^2 \quad (7)$$

$$PSNR = 10 \log_{10} \left(\frac{L^2}{MSE} \right) \quad (8)$$

$$SSIM = \frac{(2\mu_y \mu_{G(x)} + c_1)(2\sigma_y \sigma_{G(x)} + c_2)}{(\mu_y^2 + \mu_{G(x)}^2 + c_1)(\sigma_y^2 + \sigma_{G(x)}^2 + c_2)} \quad (9)$$

where y and $G(x)$ are the ground truth and synthetic images, respectively. N represents the total number of pixels in each image slice. μ_y , $\mu_{G(x)}$ and σ_y , $\sigma_{G(x)}$ are the mean and variance of the ground

truth image and the synthesis image, and $\sigma_{yG(x)}$ is the covariance between y and $G(x)$. $c_1 = (k_1L)^2$ and $c_2 = (k_2L)^2$ are two variables introduced to stabilize the SSIM division index with a weak denominator, here the default setting of $k_1 = 0.01$ and $k_2 = 0.03$ are adopted, and $L = 255$ is the dynamic range of the pixel intensity.

To evaluate the synthesis performance within the tumor region, we retrieved the clinical diagnosis report of the patients and selected twenty patients with confirmed tumors. According to the diagnosis report and tumor markers depicted by the radiologists, we manually drew three bonding boxes on the axial slice containing the tumor, i.e., the tumor region, the normal liver tissue region, and the background noise region. The tumor contrast to noise ratio (CNR) is defined as:

$$CNR = \frac{|\mu_T - \mu_N|}{\sigma_B}, \quad (10)$$

where μ_T and μ_N are the average values of the tumor region and the normal tissue region, respectively, σ_B is the standard deviation of the background ground noise region.

To evaluate the impact of image synthesis on liver tumor delineation, we invited two additional physicians (one radiation oncologist attending with >20 years of experience and one senior medical resident trained by the same attending) to independently contour 21 liver tumors from 10 testing patients. DICE coefficient (defined as intersection over average of two volumes) and Hausdorff distance (HD, measuring how far two volumes are from each other) were used as analysis metrics. Both physicians first performed tumor delineation on the real T1ce MRs. The DICE coefficients were calculated from the two physicians' contours for each tumor. The average DICE (RadOnc7 vs. RadOnc 8) served as the baselines. The attending performed delineation on both the real T1ce MRs and synthetic T1ce MRs. DICE (real vs. synth) coefficient and Hausdorff distance were calculated from volumes contoured on real and synthetic MRs for each tumor.

To further evaluate the potential effects on image-guided radiation therapy, we performed tumor center shifts analysis. In this analysis, the tumor center of mass coordinates was extracted from real and synthetic volumes. The difference in the coordinates indicated the shifts in the superior-inferior (SI), right-left (RL), and anterior-posterior (AP) directions between real and synthetic tumor volumes.

4. Results

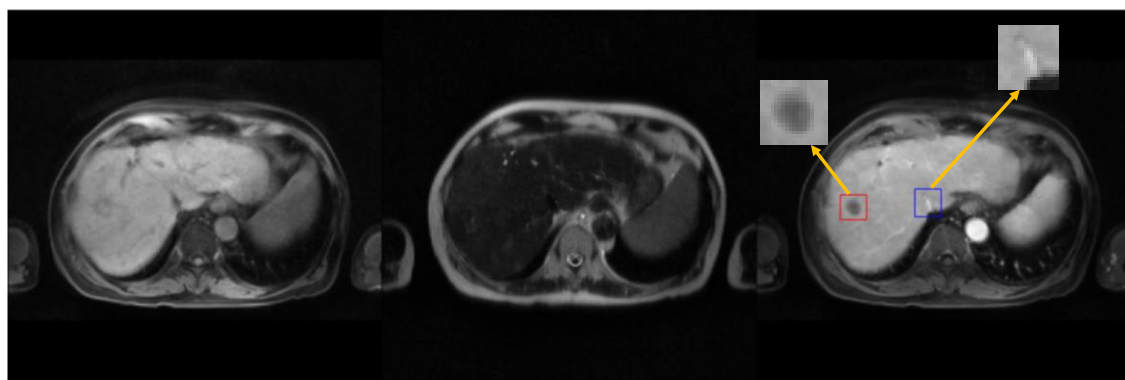
The synthesized T1ce was compared with the ground truth T1ce for thirty random patients. The overall performance is presented in Table 1. GRMM-GAN achieved a PSNR of 28.56 ± 0.87 , an SSIM of 0.869 ± 0.028 , and an MSE of 83.27 ± 15.42 , which outperformed all state-of-the-art multi-modal MR synthesis models, and the single-modal pix2pix method. GRMM-GAN significantly improved all metrics over comparison methods (p-value < 0.05). Figure 3 presents the synthesis results for one example patient. GRMM-GAN is shown to maintain rich details and textural consistency. The detail preservation is evident in the tumor region denoted by the red box.

Table 1. Performance Comparison of the Proposed GR-MMSF GAN and State-of-The-Art Comparisons in Contrast-enhanced Liver MR Synthesis, and Ablation Study of the Proposed Method.

METHODS	PSNR	SSIM	MSE
PIX2PIX (T2) [20]	24.45±1.33	0.786±0.035	213.51±29.81
PIX2PIX (T1PRE) [20]	24.82±1.42	0.795±0.039	192.32±25.69
LR-CGAN [33]	25.91±1.25	0.813±0.032	141.48±20.37
HI-NET [34]	27.28±1.26	0.836±0.036	110.86±21.04
MMGSN-NET [35]	28.04±0.93	0.851±0.033	98.43±21.16
GR-MMSF GAN (PROPOSED)	28.56±0.87	0.869±0.028	83.27±15.42
ONLY GR, NO MMD	27.61±1.06	0.838±0.031	105.43±16.15
NO GR, NO MMD	26.84±1.19	0.820±0.034	121.65±16.32

In comparison, the other models, including both the single-input Pix2pix models and multi-input fusion models (LR-cGAN, Hi-Net, and MMgSN-Net), resulted in a substantial loss of anatomical details due to the lack of complementary feature mining and discrimination. The pix2pix methods with single input could not exploit the complementary information across different MR modalities. Although the pix2pix with T1pre input, as shown in Figure 3(e), roughly predicted the tumor contour, it lost fine details, such as small vessels in the liver. The synthesis quality improved using LR-cGAN with added T2 input, as shown in Figure 3(f), but the improvement on small vessel structures was modest. Hi-Net and MMgSN-Net more substantially improved fine structural preservation at the cost of compromising the contour integrity of the hypodense liver tumors. In contrast, our method, GRMM-GAN, retained the tumor integrity and preserved the high-contrast fine features, as shown in Figure 3(i).

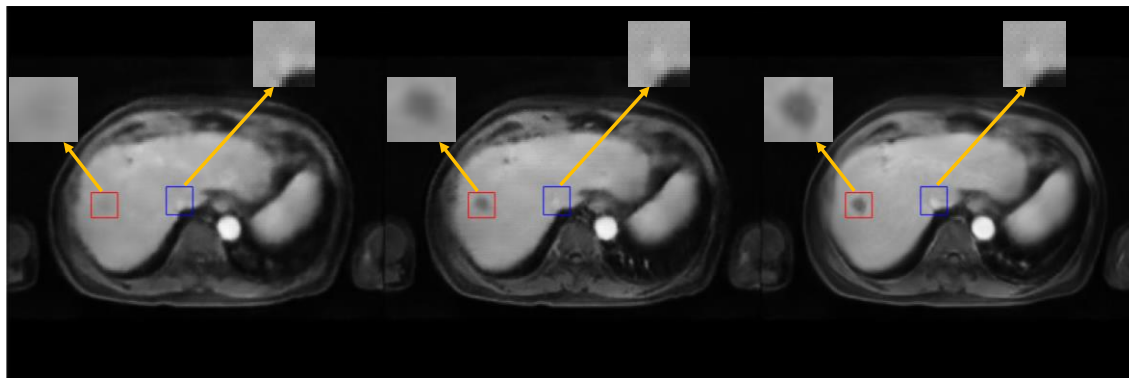
The Turing test results further substantiated the image quality preservation shown in Table 2. The average Turing test score from the six radiation oncologists was $52.33\% \pm 6.06$, close to random guessing, indicating comparable visual quality between the synthetic images and the real images.



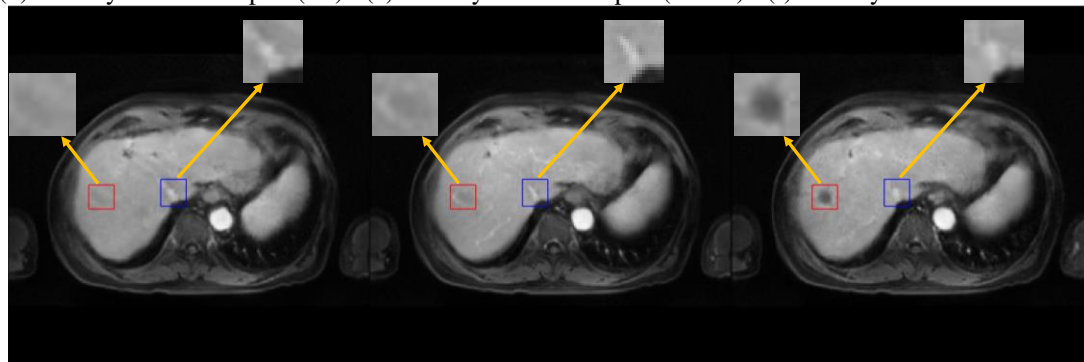
(a) Input modality, T1pre.

(b) Input modality, T2.

(c) Target image, real T1ce.



(d) T1ce synthesis Pix2pix. (T2) (e) T1ce synthesis Pix2pix. (T1Pre) (f) T1ce synthesis LR-cGAN.



(g) T1ce synthesis Hi-Net.

(h) T1ce synthesis, MMgSN-Net.

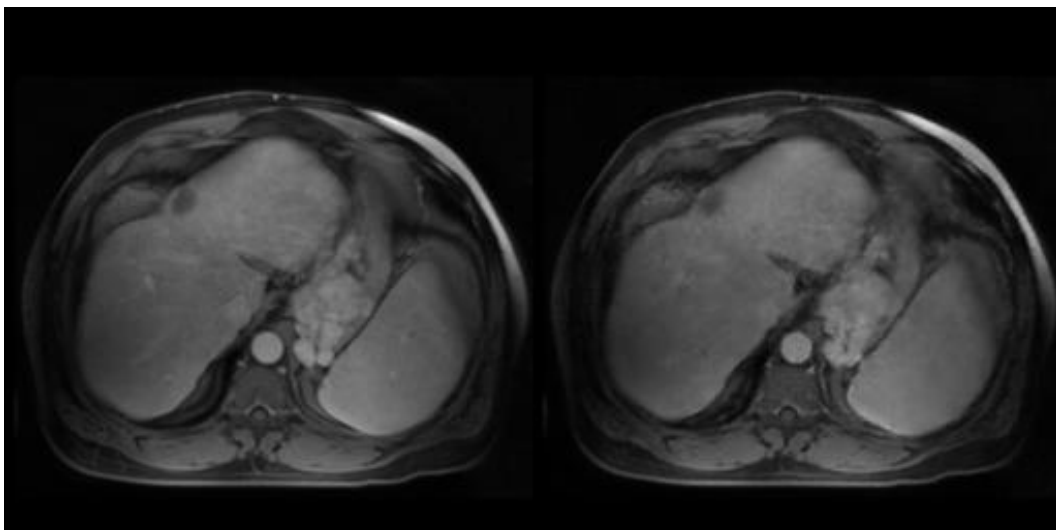
(i) T1ce synthesis, proposed

Figure 3. Synthesis example. (a) Input modality, T1pre. (b) Input modality, T2. (c) Target image, real T1ce. (d) T1ce synthesis, Pix2pix (T2). (e) T1ce synthesis, Pix2pix (T1pre). (f) T1ce synthesis, LR-cGAN. (g) T1ce synthesis, Hi-Net. (h) T1ce synthesis, MMgSN-Net. (i) T1ce synthesis, GRMM-GAN.

Table 2. Turing Test Results from Six Radiation Oncologists.

Radiation Oncologist	Evaluation	Results	Percentage (correct)
1	Correct	55	55%
	Incorrect	45	
2	Correct	56	56%
	Incorrect	44	
3	Correct	42	42%
	Incorrect	58	
4	Correct	59	59%
	Incorrect	41	
5	Correct	53	43%
	Incorrect	47	
6	Correct	49	49%
	Incorrect	51	
Average			52.3%

The contribution of each module was evaluated in an ablation test. The performance by different statistics metrics, PSNR, MSE, and SSIM, showed a consistent trend of additive value. Therefore, only MSE is described and discussed here for brevity. Figure 4 provides a visual evaluation of the ablation study. After removing the multi-scale and multi-smoothness (MMD) component, the average performance indicated by MSE increased from 83.27 ± 15.42 to 105.43 ± 16.15 , highlighting the contribution of MMD in the discrimination of the low-quality synthesis. The exclusion of the GR model further increased the MSE to 121.65 ± 16.32 , indicating the contribution of GR to structural preservation. Additionally, the average PSNR, SSIM, and MSE for the synthesized tumor region are 28.40, 0.856, and 88.71, respectively. The multiple performance evaluations show that the real and synthesized MR images for the specific tumor region are very similar, leading to the conclusion that the synthesized MR images could be a suitable surrogate for real contrast MR when the latter is unobtainable.



(a) Target image, real T1ce.

(b) GR+MMD

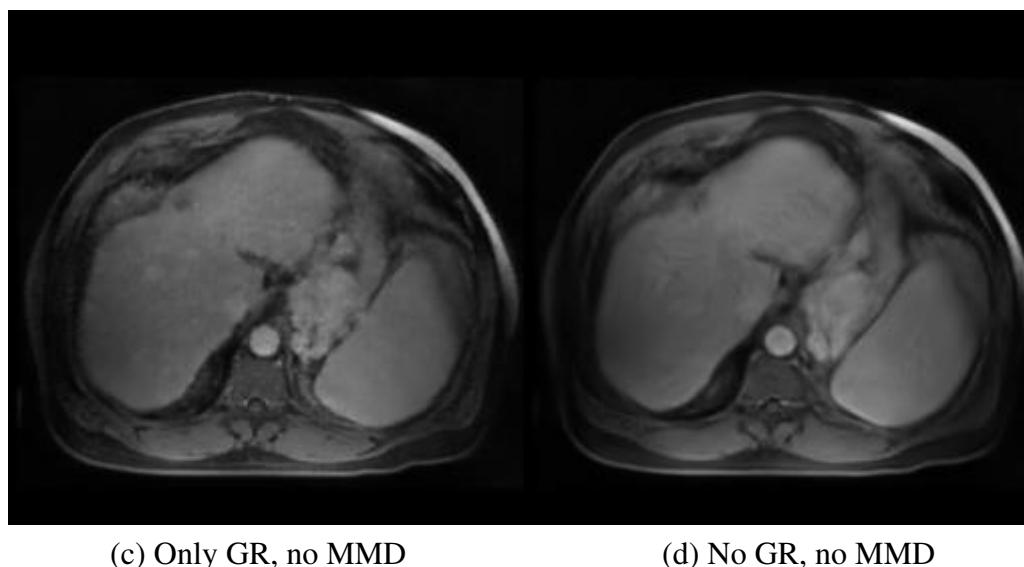


Figure 4. Visual examples for ablation study. (a) Target image, real T1ce. (b) GR+MMD. (c) Only GR, no MMD. (d) No GR, no MMD.

The average CNR of twenty patients for the real MR tumor region was 26.18 ± 21.13 . In contrast, the average CNR for the synthesized MR tumor regions was 24.53 ± 20.08 with a p-value of 0.401, with no significant difference between the real and synthesized images for tumor analysis, suggesting similar tumor conspicuity provided by synthetic and real T1ce MR images.

RadOnc 7 and 8 achieved an average DICE (RadOnc7 vs. RadOnc 8) of 0.91 ± 0.02 , from tumor volumes drawn on the real T1ce MRs. This result sets the inter-operator baseline in the real clinical setting. RadOnc 8 achieved an average DICE (real vs. synth) of 0.90 ± 0.04 and HD of 4.76 ± 1.82 mm. Only sub-millimeter tumor center shifts were observed in all three directions. The detailed tumor volume information is shown in Table 3.

Table 3. Tumor contouring evaluation on ten patients (with 21 lesions) from two radiation oncologists.

Radiation Oncologist	Average (range) volume (cc)	SI (mm)	RL (mm)	AP (mm)
7 (on real T1ce)	30.8 (1.2-233.7)	NA		
8 (on real T1ce)	29.4 (1.1-238.4)	0.67	0.41	0.39
8 (on synthetic T1ce)	28.6 (1.1-245.0)			

5. Discussion

Contrast-enhanced MR images provide improved visualization of the liver tumor, which is essential for MR-guided radiotherapy and SBRT. However, repeated injections of contrast may not be clinically viable. As an alternative, we synthesized contrast-enhanced MR images from multi-modality pre-contrast MR images using a novel GRMM-GAN model. The synthesized virtual contrast-enhanced MR images closely mimicked the ground truth contrast-enhanced images in quantitative image analysis and human expert Turing tests. The success indicates that the pre-contrast T1 and T2 images have a substantial predictive value for the post-contrast MR. The latent information in T1 and T2 images is intricate for human operators to appreciate but can be distilled via image synthesis into a format, e.g., post-contrast MR, familiar to human operators. Our technical innovations in building the GAN network overcame multiple challenges in existing image synthesis methods.

Several deep learning models have been developed for brain MR synthesis [34,48]. Compared with brain images, abdominal MR images are considerably more heterogeneous in soft tissue composition, MR relaxation properties, size, shape, and textures. The heterogeneity is further

compounded by substantial motion in this anatomical region, leading to mismatches among different MR sequence acquisitions. Therefore, abdominal MR synthesis is a more challenging problem.

The challenge is highlighted by prior efforts at using single-modal deep learning models for abdominal MR image synthesis [49,50], which fail to learn the salient knowledge across MR modalities, leading to a substantial loss in the synthesized image quality. The proposed GRMM-GAN is the first network to discover and fuse mutual-complementary information, which markedly improves the synthesis performance, showing realistic contrast enhancement style translation, precise contours, and textures.

Our technical innovations are summarized as follows. First, a multi-modal fusion model was developed to exploit the salient and complementary information in multiple MR modalities. A sparse attention fusion module was investigated to distinguish the redundancy incurred. Second, we adopted the previously proposed image gradient regularization mechanism to avoid the loss of anatomical details in the GAN-based model and presented the novel multi-scale and multi-smoothness discrimination. In addition to these technical innovations, we prepared a valuable multi-modal abdominal database for contrast enhancement MR synthesis with patient data (165 scans from 61 patients) to support model development and validation.

The exclusion of the GR and MMD components decreases the average MSE to 105.43 ± 16.15 (P-value < 0.05) and 121.65 ± 16.32 (P-value < 0.05), respectively. Figure 4 provides an intuitive understanding of the essential contribution of each module. The absence of the MMD module degrades the discriminative ability of the network in over-smoothing and blurring effects, which can be visually observed in Figure 4 (c). The further removal of the GR module decreases the perception of the network to textural and structural details, as illustrated in Figure 4 (d). The specific tumor region synthesis evaluation also solidifies the proposed method in potential clinical application. The CNR and other performance evaluation metrics (PSRN, SSIM, and MSE) indicate that the salience and similarity of the synthesized specific tumor of interest are good enough compared with the real tumor data.

Similar to our GRMM-GAN model, the three multi-modal fusion generative methods, LR-cGAN, Hi-Net, and MMgSN-Net, generally performed better than the single modality synthesis model (pix2pix). The better performance could be attributed to the theory of multi-modal fusion to exploit the complementary information from different MR modalities. Specifically, the MMgSN-Net was derived from Hi-Net and achieved the second-best overall performance boosted by the self-attention fusion module multi-modal feature learning. However, these methods still failed to discriminate the redundancy that existed accordingly, resulting in inferior synthesis quality with over-smoothing effects. For the single modal synthesis comparison, the pix2pix model yielded better performance with the T1pre as input than with T2, likely because T1pre is more similar to the output T1ce modality. On the contrary, Li et al. [35] reported that the crucial knowledge from multi-modalities (e.g., T2) in revealing the hyper-intensity or pathological inhomogeneity characteristics could not be neglected.

Despite the excellent image synthesis quality and expert evaluation results, our work has the following limitations. To balance the image quality and the computation complexity, we downsampled the original MR images from 256×256 . This preprocessing could compromise the fidelity and precision of the synthesis results compared to the original resolution. In the future, a super-resolution [43,51] or interpolation [52,53] could be incorporated. Although quantitative metrics, such as PSNR, MSE, and SSIM, are suited for comparative image quality tests, they are not directly linked to more clinically relevant dosimetric measures for adaptive planning applications. We have performed an exploratory contouring study and showed satisfactory DICE coefficients calculated from real vs. synthetic MRs. In practice, the impact on the dosimetric endpoints will need to be evaluated for adaptive radiation therapy, which will require longitudinal MR images through the treatment course. Finally, our model is built upon diagnostic MR images; whether the network can be applied to MR-LINAC systems during, or post-radiation MR synthesis will need to be tested.

6. Conclusion

This study developed a novel multi-modal contrast-enhanced liver MR image synthesis network (GRMM-GAN) capable of synthesizing highly realistic T1ce images that outperformed the three state-of-the-art synthesis networks.

Supplementary Materials:

Funding: This work is supported by NIH R01EB029088.

Research data are stored in an institutional repository and will be shared with the corresponding author upon request.

Conflicts of Interest: None.

References

1. Key Statistics About Liver Cancer. Link: <https://www.cancer.org/cancer/liver-cancer/about/what-is-key-statistics.html>.
2. Chartampilas E, Rafailidis V, Georgopoulou V, et al. Current Imaging Diagnosis of Hepatocellular Carcinoma. *Cancers (Basel)*. 2022;14:3997.
3. Zhao D, Cao J, Zhang L, et al. Targeted Molecular Imaging Probes Based on Magnetic Resonance Imaging for Hepatocellular Carcinoma Diagnosis and Treatment. *Biosensors (Basel)*. 2022;12:342.
4. European Association for the Study of the Liver. EASL 2017 Clinical Practice Guidelines on the management of hepatitis B virus infection. *J Hepatol*. 2017;67:370–398.
5. Gordan JD, Kennedy EB, Abou-Alfa GK, et al. Systemic Therapy for Advanced Hepatocellular Carcinoma: ASCO Guideline. *J Clin Oncol*. 2020;38:4317–4345.
6. Heimbach JK, Kulik LM, Finn RS, et al. AASLD guidelines for the treatment of hepatocellular carcinoma. *Hepatology*. 2018;67:358–380.
7. Zhang X-F, Lai L, Zhou H, et al. Stereotactic body radiotherapy plus transcatheter arterial chemoembolization for inoperable hepatocellular carcinoma patients with portal vein tumour thrombus: A meta-analysis. *PLoS One*. 2022;17:e0268779.
8. Lewis S, Barry A, Hawkins MA. Hypofractionation in Hepatocellular Carcinoma - The Effect of Fractionation Size. *Clin Oncol (R Coll Radiol)*. 2022;34:e195–e209.
9. Velec M, Moseley JL, Svensson S, et al. validation of biomechanical deformable image registration in the abdomen, thorax, and pelvis in a commercial radiotherapy treatment planning system. *Med Phys*. 2017;44:3407–3417.
10. Sadowski EA, Bennett LK, Chan MR, et al. Nephrogenic systemic fibrosis: risk factors and incidence estimation. *Radiology*. 2007;243:148–157.
11. Wahsner J, Gale EM, Rodríguez-Rodríguez A, et al. Chemistry of MRI contrast agents: current challenges and new frontiers. *Chemical Reviews*. 2018;119:957–1057.
12. Kleesiek J, Morshuis JN, Isensee F, et al. Can virtual contrast enhancement in brain MRI replace gadolinium?: a feasibility study. *Investigative Radiology*. 2019;54:653–660.
13. Slack A, Yeoman A, Wendon J. Renal dysfunction in chronic liver disease. *Crit Care*. 2010;14:214.
14. Marckmann P, Skov L, Rossen K, et al. Nephrogenic systemic fibrosis: suspected causative role of gadodiamide used for contrast-enhanced magnetic resonance imaging. *Journal of the American Society of Nephrology*. 2006;17:2359–2362.
15. Thomsen HS, Morcos SK, Dawson P. Is there a causal relation between the administration of gadolinium based contrast media and the development of nephrogenic systemic fibrosis (NSF)? *Clinical Radiology*. 2006;61:905–906.
16. Montalt-Tordera J, Quail M, Steeden JA, et al. Reducing Contrast Agent Dose in Cardiovascular MR Angiography with Deep Learning. *Journal of Magnetic Resonance Imaging*. 2021;54:795–805.
17. Lee G, Fujita H eds. *Deep Learning in Medical Image Analysis: Challenges and Applications*. Cham: Springer International Publishing; 2020.
18. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. *Advances in Neural Information Processing Systems*. 2014;27:1–9.
19. Mirza M, Osindero S. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*. 2014.
20. Isola P, Zhu J-Y, Zhou T, et al. Image-to-image translation with conditional adversarial networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.; 2017:1125–1134.
21. Zhu J-Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: *Proceedings of the IEEE International Conference on Computer Vision*.; 2017:2223–2232.

22. Wolterink JM, Leiner T, Viergever MA, et al. Generative adversarial networks for noise reduction in low-dose CT. *IEEE Transactions on Medical Imaging*. 2017;36:2536–2545.
23. Huang Z, Liu X, Wang R, et al. Considering anatomical prior information for low-dose CT image enhancement using attribute-augmented wasserstein generative adversarial networks. *Neurocomputing*. 2021;428:104–115.
24. Bai T, Du M, Zhang L, et al. A novel Alzheimer's disease detection approach using GAN-based brain slice image enhancement. *Neurocomputing*. 2022;492:353–369.
25. Li Y, Iwamoto Y, Lin L, et al. VolumeNet: a lightweight parallel network for super-resolution of MR and CT volumetric data. *IEEE Transactions on Image Processing*. 2021;30:4840–4854.
26. You C, Li G, Zhang Y, et al. CT super-resolution GAN constrained by the identical, residual, and cycle learning ensemble (GAN-CIRCLE). *IEEE Transactions on Medical Imaging*. 2019;39:188–203.
27. Nie D, Trullo R, Lian J, et al. Medical image synthesis with deep convolutional adversarial networks. *IEEE Transactions on Biomedical Engineering*. 2018;65:2720–2730.
28. Boni KNB, Klein J, Vanquin L, et al. MR to CT synthesis with multicenter data in the pelvic area using a conditional generative adversarial network. *Physics in Medicine & Biology*. 2020;65:075002.
29. Dar SU, Yurt M, Karacan L, et al. Image synthesis in multi-contrast MRI with conditional generative adversarial networks. *IEEE Transactions on Medical Imaging*. 2019;38:2375–2388.
30. Kim S, Jang H, Hong S, et al. Fat-saturated image generation from multi-contrast MRIs using generative adversarial networks with Bloch equation-based autoencoder regularization. *Medical Image Analysis*. 2021;73:102198.
31. Denck J, Guehring J, Maier A, et al. MR-contrast-aware image-to-image translations with generative adversarial networks. *International Journal of Computer Assisted Radiology and Surgery*. 2021;16:2069–2078.
32. Chartsias A, Joyce T, Giuffrida MV, et al. Multi-modal MR synthesis via modality-invariant latent representation. *IEEE Transactions on Medical Imaging*. 2017;37:803–814.
33. Zhan B, Li D, Wang Y, et al. LR-cGAN: Latent representation based conditional generative adversarial network for multi-modality MRI synthesis. *Biomedical Signal Processing and Control*. 2021;66:102457.
34. Zhou T, Fu H, Chen G, et al. Hi-Net: hybrid-fusion network for multi-modal MR image synthesis. *IEEE Transactions on Medical Imaging*. 2020;39:2772–2781.
35. Li W, Xiao H, Li T, et al. Virtual Contrast-Enhanced Magnetic Resonance Images Synthesis for Patients With Nasopharyngeal Carcinoma Using Multimodality-Guided Synergistic Neural Network. *International Journal of Radiation Oncology* Biology* Physics*. 2022;112:1033–1044.
36. Yurt M, Dar SU, Erdem A, et al. mustGAN: multi-stream generative adversarial networks for MR image synthesis. *Medical Image Analysis*. 2021;70:101944.
37. Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution. In: *European Conference on Computer Vision*. Springer; 2016:694–711.
38. Jin X, Chen Z, Li W. AI-GAN: Asynchronous interactive generative adversarial network for single image rain removal. *Pattern Recognition*. 2020;100:107143.
39. Jiao C, Chen C, Gou S, et al. L₁ Sparsity-Regularized Attention Multiple-Instance Network for Hyperspectral Target Detection. *IEEE Transactions on Cybernetics*. 2021:1–14.
40. Gou S, Liu W, Jiao C, et al. Gradient regularized convolutional neural networks for low-dose CT image enhancement. *Physics in Medicine & Biology*. 2019;64:165017.
41. Sobel I, Feldman G. An isotropic 3x3 image gradient operator. *Presentation at Stanford AI Project*. 1968.
42. Yu B, Zhou L, Wang L, et al. Ea-GANs: Edge-Aware Generative Adversarial Networks for Cross-Modality MR Image Synthesis. *IEEE Transactions on Medical Imaging*. 2019;38:1750–1762.
43. Wang T-C, Liu M-Y, Zhu J-Y, et al. High-Resolution Image Synthesis and Semantic Manipulation With Conditional GANs. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.; 2018:8798–8807.
44. Durugkar I, Gemp I, Mahadevan S. Generative Multi-Adversarial Networks. *arXiv preprint arXiv:1611.01673*. 2017.
45. Lowe DG. Object recognition from local scale-invariant features. In: *Proceedings of the Seventh IEEE International Conference on Computer Vision*. Vol 2.; 1999:1150–1157 vol.2.
46. Lowe DG. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*. 2004;60:91–110.
47. Turing AM. Computing Machinery and Intelligence. In: Epstein R, Roberts G, Beber G, eds. *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer*. Dordrecht: Springer Netherlands; 2009:23–65.
48. Xie H, Lei Y, Wang T, et al. Magnetic resonance imaging contrast enhancement synthesis using cascade networks with local supervision. *Medical Physics*. 2022;49:3278–3287.

49. Zhao J, Li D, Kassam Z, *et al.* Tripartite-GAN: synthesizing liver contrast-enhanced MRI to improve tumor detection. *Medical Image Analysis*. 2020;63:101667.
50. Xu C, Zhang D, Chong J, *et al.* synthesis of gadolinium-enhanced liver tumors on nonenhanced liver MR images using pixel-level graph reinforcement learning. *Medical Image Analysis*. 2021;69:101976.
51. Zhao X, Zhang Y, Zhang T, *et al.* Channel Splitting Network for Single MR Image Super-Resolution. *IEEE Transactions on Image Processing*. 2019;28:5649–5662.
52. Keys R. Cubic convolution interpolation for digital image processing. *IEEE Transactions on Acoustics, Speech, and Signal Processing*. 1981;29:1153–1160.
53. Valenberg W, Klein S, Vos FM, *et al.* An Efficient Method for Multi-Parameter Mapping in Quantitative MRI Using B-Spline Interpolation. *IEEE Transactions on Medical Imaging*. 2020;39:1681–1689.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.