Radiotherapy and Oncology 150 (2020) 217–224

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Original Article

Magnetic resonance-based synthetic computed tomography images generated using generative adversarial networks for nasopharyngeal carcinoma radiotherapy treatment planning

Yinglin Peng [a,b,1], Shupeng Chen [c,*,1], An Qin [c], Meining Chen [a], Xingwang Gao [d], Yimei Liu [a], Jingjing Miao [a], Huikuan Gu [a], Chong Zhao [a], Xiaowu Deng [a], Zhenyu Qi [a,*]

[a] Department of Radiation Oncology, Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine; [b] School of Biomedical Engineering, Sun Yat-sen University, Guangzhou, China; [c] Department of Radiation Oncology, William Beaumont Hospital, Royal Oak, United States; [d] Department of Radiation Oncology, Guangdong Provincial People's Hospital, Guangzhou, China

ARTICLE INFO

Article history:
Received 13 December 2019
Received in revised form 23 June 2020
Accepted 25 June 2020
Available online 3 July 2020

Keywords:
Synthetic CT
Nasopharyngeal carcinoma
MRI-only radiotherapy
Generative adversarial networks
Conditional GAN
Cycle GAN

ABSTRACT

Background and purpose: To investigate the feasibility of synthesizing computed tomography (CT) images from magnetic resonance (MR) images using generative adversarial networks (GANs) for nasopharyngeal carcinoma (NPC) intensity-modulated radiotherapy (IMRT) planning.

Materials and methods: Conventional T1-weighted MR images and CT images were acquired from 173 NPC patients. The MR and CT images of 28 patients were randomly chosen as the independent tested set. The remaining images were used to build a conditional GAN (cGAN) and a cycle-consistency GAN (cycleGAN). A U-net was used as the generator in cGAN, whereas a residual-Unet was used as the generator in cycleGAN. The cGAN was trained using the deformable registered MR-CT image pairs, whereas the cycleGAN was trained using the unregistered MR and CT images. The generated synthetic CT (SCT) images from cGAN and cycleGAN were compared with the true CT images with respect to their Hounsfield Unit (HU) discrepancy and dosimetric accuracy for NPC IMRT plans.

Results: The mean absolute errors within the body were 69.67 ± 9.27 HU and 100.62 ± 7.39 HU for the cGAN and cycleGAN, respectively. The 2%/2-mm γ passing rates were (98.68 ± 0.94)% and (98.52 ± 1.13)% for the cGAN and cycleGAN, respectively. Meanwhile, the absolute dose discrepancies within the regions of interest were (0.49 ± 0.24)% and (0.62 ± 0.36)%, respectively.

Conclusion: Both cGAN and cycleGAN could swiftly generate accurate SCT volume images from MR images, with high dosimetric accuracy for NPC IMRT planning. cGAN was preferable if high-quality MR-CT image pairs were available.

Magnetic resonance (MR) imaging is widely used in radiotherapy to better delineate the tumor target and organs at risk owing to its superior soft-tissue contrast compared with that of computed tomography (CT) imaging [1]. Developing a novel radiotherapy modality based solely on MR simulation, i.e., MR-only radiotherapy, is becoming a common research focus [2], because this eliminates extra CT imaging and thus reduces patient radiation exposure. However, implementing MR-only radiotherapy requires deriving synthetic CT (SCT) images from MR images to provide electron densities for radiation dose calculation. Accordingly, various methods have been proposed for deriving SCT images, including statistical modeling [3], traditional machine learning [4,5], and multi-atlas based methods [6–14]. Recently, increasing interest has been focused on artificial intelligence-based methods, such as deep convolutional neural network (CNN) [15,16,17–25], conditional generative adversarial network (cGAN) [26–30], and cycle-consistent generative adversarial network (cycleGAN) [31–37].

The deep CNN model is more capable of learning a complex MR-to-CT mapping from large-scale image samples than the previous methods. Additionally, it achieves faster model deployment with only a few seconds to generate an SCT volume image once the model has been trained [38]. However, such a training a deep CNN model can cause image blurriness because the L_1 or L_2 norm is minimized by averaging all plausible outputs [39]. To reduce

* Corresponding authors at: Department of Radiation Oncology, William Beaumont Hospital, Royal Oak, MI, United States (S. Chen). Department of Radiation Oncology, Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou, China (Z. Qi).

E-mail addresses: shupeng.chen@beaumont.org (S. Chen), qizhy@sysucc.org.cn (Z. Qi).

[1] Yinglin Peng and Shupeng Chen contributed equally to this study.

https://doi.org/10.1016/j.radonc.2020.06.049
0167-8140/

the image blurriness, cGAN added an adversarial loss [40] term in the loss function to force the generated image to be indistinguishable from the real images [41]. From this perspective, cGAN is more suitable for accurate MR-CT transformation. However, training a deep CNN or cGAN relies on perfectively registered MR-CT images, which are sometimes challenging to realize. CycleGAN [42] addresses this limitation by training two mappings, i.e., MR-to-CT and CT-to-MR, simultaneously based on the principle of cycle consistency. Thus, cycleGAN has attracted great interest because it can be trained without requiring high-quality MR-CT paired images.

Deep learning models have been applied for generating SCT images in radiotherapy planning for the prostate or pelvic [18,27,43], brain [17,20,23,25,30], liver [44–46], breast [29], and head and neck cancer [21,24,47–49]. However, for cancers with more complicated treatment sites such as nasopharyngeal carcinoma (NPC), generating SCT images from conventional MR image remains challenging because several air-bearing bone structures near and/or within the nasopharynx. Tissues such as bone and air exhibit a wide range of CT numbers, but both appear low signals on conventional MR images. This may lead to great uncertainty in SCT generation and result in dose calculation inaccuracy.

Previous studies mainly trained the deep learning models using a relatively small dataset. In this study, we collected MR and CT images from a large cohort of patients. Two GAN models, i.e., cGAN and cycleGAN, were built for generating SCT image. The cGAN was trained with deformable registered MR-CT image pairs, whereas the cycleGAN was trained with unregistered MR and CT images. The model performances were assessed for different number of training samples. The dosimetric accuracy was evaluated on NPC IMRT plans.

Materials and methods

Data acquisition and preprocessing

MR and CT images at the radiotherapy position were acquired within an interval of 0.5–2 h for 173 patients with NPC. The MR images were acquired using a 3T Ingenia high-field system (Philips Healthcare) with a vendor-supplied phased-array dStream Head-Neck-Spine coil and a dedicated radiotherapy flat couch. A T1-weighted fast-field echo was used with an echo time/repetitive time of 4.9/2.4 ms, a flip angle of 12°, and a reconstructed voxel size of $0.74\times0.74\times3$ mm^3. The CT images were acquired using a SOMATOM Definition AS CT simulator (Siemens Medical System) with a reconstructed voxel size of $0.97\times0.97\times3$ mm^3. The intensity of each MR volume image was normalized as zero mean and unit variance and then scaled to a similar numeric range. The details of the intensity normalization are described in [Supplement MR intensity normalization].

A binary body mask was applied to override voxels outside the patient body for each image. To create better matched MR-CT image pairs, we performed deformable image registration between MR and CT images for each pair using an inverse-consistent diffeomorphic image registration algorithm [50]. The resulting paired MR-CT images shared a resolution of $0.74\times0.74\times3$ mm^3 with 384×384 voxels per axial slice. The deformable registered MR-CT image pairs were used to train the cGAN, whereas the unregistered MR and CT images were used to train the cycleGAN.

Network architectures

Compared with CNN, cGAN can generate a more realistic SCT image with less image blurriness. cGAN contains two CNNs, as shown in [Fig. 1(a)], i.e., one generator and one discriminator. The generator (G) is a nonlinear function with an input of MR image slice and output of the corresponding SCT image slice. The U-net architecture was used in the generator because it is capable of high-spatial-resolution prediction tasks. Unlike the original U-net [51], instance normalization was applied immediately after each convolutional layer except the last layer. Compared with batch normalization [52], instance normalization can better preserve image details during image generation, as demonstrated previously [53]. All convolution filters in the generator measured 3×3, except for the last layer that measured 1×1. We applied zero padding for each convolutional layer to preserve the size of output feature maps after applying convolution operations. The discriminator (D) is a classifier to classify a CT image being real or synthetic. It contains 6 successively convolution layers with the filter number of 64, 128, 256, 512, 512, and 1 for each layer. All filters in the discriminator were 4×4 with a stride of 2 except the last layer, which uses a stride of 1. These hyper-parameters, such as strides and filter number/size, were initially selected based on previous GAN studies [29,49] and were then fine-tuned considering the GPU memory, the computational speed, and the stability of the training process. Instance normalization was applied after each convolutional layer and before the activation in the discriminator. The LeakyReLU [54] with a slope parameter 0.3 was used for all activations.

Compared with CNN and cGAN, cycleGAN can be trained with unpaired MR and CT images. cycleGAN contains four CNNs, as shown in [Fig. 1(b)], i.e., two generators and two discriminators. The first generator ($G_{MR\text{-}CT}$) takes an MR slice as input and generates the SCT slice, and the first discriminator (D_{CT}) classifies a CT image being real or synthetic. The second generator ($G_{CT\text{-}MR}$) takes a CT slice as input and generates the synthetic MR slice, and the second discriminator (D_{MR}) classifies an MR image being real or synthetic. A hybrid network architecture res-Unet was developed and used in the generators. It contains 9 residual blocks with max-pooling/unpooling layers. A schematic view of the res-Unet architecture is depicted in [Fig. 1(c)]. The hybrid network architecture has the advantages of both U-net [51] and residual network [55]. It can learn the local features effectively through the long skip connections like U-net. The deep convolutional layers can be effectively trained with the residual connections to alleviate the gradient vanishing issue. These advantages have been demonstrated by a previous study [32]. In this study, the network architecture of discriminator used in cycleGAN was the same as that used in cGAN.

Network training

The cGAN was trained with deformable registered MR-CT image pairs. Given an image pair of $\{MR_i, CT_i\}$, $i = 1, 2\ldots N$ (N is the total number of paired image slices), the training process involved identifying the optimal parameters θ_G and θ_D for the generator (G) and the discriminator (D) respectively, which was denoted as:

$$\theta_G = arg\min_{\theta_D}\left(|G(MR_i) - CT_i| + \lambda\cdot(D(G(MR_i)))^2\right) \quad (1)$$

and

$$\theta_D = arg\max_{\theta_G}(D(G(MR_i), CT_i))^2 \quad (2)$$

where λ is the weighting factor empirically chosen to be 5. Specifically, it was initially selected based on previous studies on cGAN [27,47,49], and was then fine-tuned to balance the image contrast and the MAE loss term. We used the least square term, instead of cross-entropy used in the original cGAN [41], in the loss functions because it can improve the stabilization of GAN training [56]. For an iteration of training, the θ_G was updated once and θ_D was updated twice successively based on the randomly selected paired

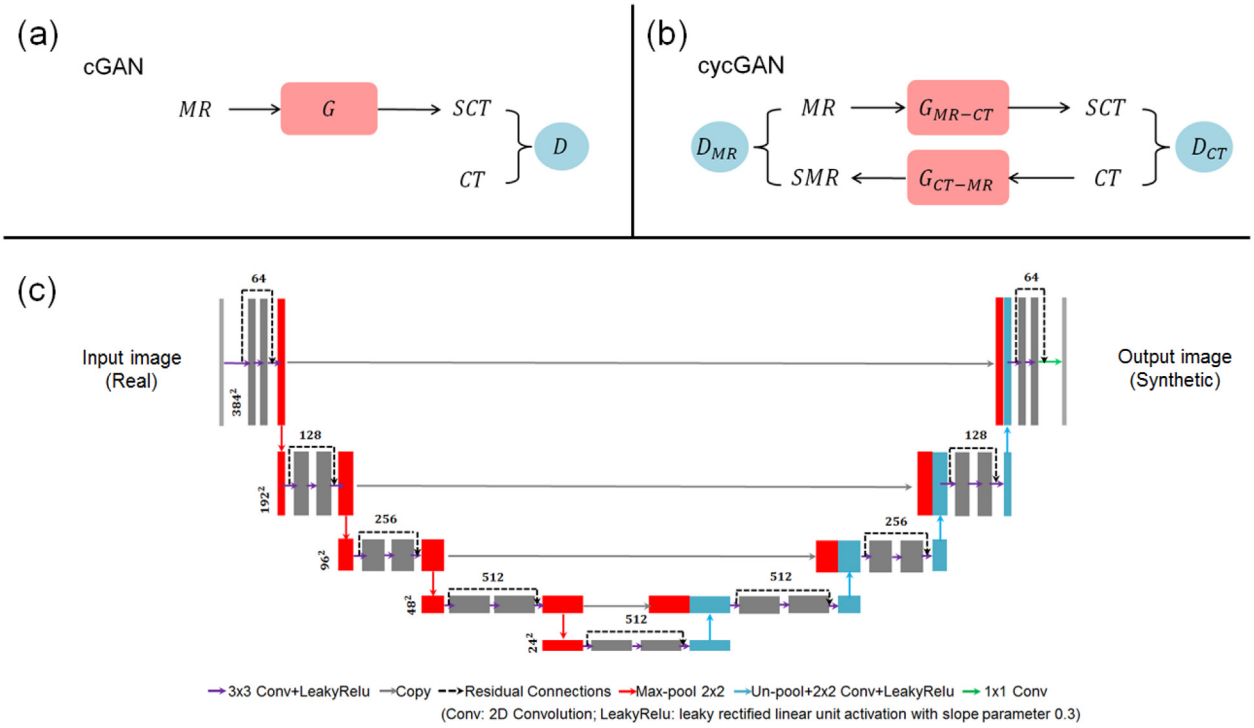

Fig. 1. Illustration of the cGAN (a), cycleGAN (b), and residual-Unet architecture used in the cycleGAN (c). G and D indicate the generator and discriminator used in cGAN, respectively; G_{MR-CT} and G_{CT-MR}, generators; and D_{MR} and D_{CT}, discriminators used in cycleGAN. *Abbreviations*: cGAN = conditional generative adversarial network; cycleGAN = cycle-consistent generative adversarial network; SCT = synthetic CT image; SMR = synthetic MR image.

image slices $\{MR_i, CT_i\}$. The θ_D was fixed while training G and the θ_G was fixed while training D.

The cycleGAN was trained with unregistered MR and CT images based on the principle of cycle-consistency [42], which was denoted as: $G_{MR\text{-}CT}(G_{CT\text{-}MR}(CT)) \approx CT$ and $G_{CT\text{-}MR}(G_{MR\text{-}CT}(MR)) \approx MR$, where $G_{MR\text{-}CT}$ and $G_{CT\text{-}MR}$ are the generators from MR to CT and from CT to MR, respectively. Given unpaired images MR_i, $i = 1, 2 \ldots N$ and CT_j, $j = 1, 2 \ldots M$ (N and M are the total numbers of unregistered MR and CT image slices), the training process was to identify the optimal parameters $\theta_{G_{MR-CT}}$, $\theta_{G_{CT-MR}}$, $\theta_{D_{CT}}$, and $\theta_{D_{MR}}$ for the generators ($G_{CT\text{-}MR}$ and $G_{MR\text{-}CT}$) and discriminators (D_{CT} and D_{MR}), which were denoted as:

$$\begin{cases} \theta_{G_{MR-CT}} = \arg\min_{\theta_{D_{MR}}} \left(|G_{MR-CT}(G_{CT-MR}(CT_i)) - CT_i| + \lambda_1 \cdot (D_{MR}(G_{CT-MR}(CT_i)))^2 \right) \\ \theta_{G_{CT-MR}} = \arg\min_{\theta_{D_{CT}}} \left(|G_{CT-MR}(G_{MR-CT}(MR_j)) - MR_j| + \lambda_2 \cdot (D_{CT}(G_{MR-CT}(MR_j)))^2 \right) \end{cases} \quad (3)$$

and

$$\begin{cases} \theta_{D_{CT}} = \arg\max_{\theta_{G_{MR-CT}}} \left(D_{CT}(G_{MR-CT}(MR_j), CT_i) \right)^2 \\ \theta_{D_{MR}} = \arg\max_{\theta_{G_{CT-MR}}} \left(D_{MR}(G_{CT-MR}(CT_i), MR_j) \right)^2 \end{cases} \quad (4)$$

where λ_1 and λ_2 are the weighting factors and were chosen to be 10, which was determined with the "trial and error" method to balance the image contrast and the MAE loss term, based on previous studies on MR-based SCT image generation using cycleGAN models [32,34]. For an iteration of the training process, the parameters of the generators, $\theta_{G_{MR-CT}}$ and $\theta_{G_{CT-MR}}$, were updated once and the parameters of the discriminators, $\theta_{D_{CT}}$ and $\theta_{D_{MR}}$, were updated twice successively based on the randomly selected image slices, MR_i and CT_j.

Implementation details

We used a mini-batch of 3 and 1 for training the cGAN and cycleGAN, respectively. To alleviate overfitting, we performed data augmentation by applying deformations to the samples during training. Data augmentation is described in [Supplementary Data augmentation details]. The models were trained on an NVIDIA GP100 GPU with 16 GB memory and an M6000 GPU with 12 GB memory. The Initial learning rates for the discriminators and the generators were 0.0001 and 0.0002, respectively, and decreased to half when the training error plateaued. The training process was stopped when the validation error started increasing. Adam optimization was used for training with the momentum hyper-parameters β_1 and β_2 set to be 0.5 and 0.9, respectively. All models were trained from scratch and implemented using *Keras* (v2.2.4) with *Tensorflow* (v1.13.1) as the backend.

Evaluation

In total, 28 of the 173 MR-CT image pairs were randomly selected as the independent tested set, and the other 145 pairs were used to build the models, including 135 pairs as the training set and 10 pairs as the validation set. For each tested patient, the MR image was used as inputs for the cGAN and the cycleGAN to generate the SCT images. The paired CT image was referred to as a true CT image in this study. The mean error (ME) and mean absolute error (MAE) between the SCT images and the true CT images were used to measure the CT number estimation accuracy:

$$ME = \frac{1}{N}\sum_{i}^{N}(SCT(v_i) - CT(v_i)) \quad (5)$$

and

$$MAE = \frac{1}{N}\sum_{i}^{N}|SCT(v_i) - CT(v_i)| \quad (6)$$

where v is a voxel within regions of interest, and N is the total number of voxels. The ME and MAE were calculated within the patient body, bone (Hounsfield Unit, HU $\geq$ 150), soft tissue (150 > HU $\geq$ −100), and air (HU < −100), respectively. The ME and MAE of the SCT images generated by cGAN were compared with those of cycleGAN. To further investigate the effect of training data size on model performance, we also trained the models by reducing the training set to 68 (50%) and 34 (25%) pairs. Meanwhile, the tested set and validation set remained unchanged. The generated SCT images were compared with the true CT images.

The dosimetric accuracy of using the SCT images for planning was analyzed using clinical NPC treatment plans. SCT images generated by the models trained with all samples (from the 135 patients) were included in the dosimetric evaluation. A dose of 7000 Gy was prescribed to the primary tumor target. For each tested patient, an IMRT plan with 9 beams of 6-MV X-rays was designed using the true CT images (Monaco, v5.1, Elekta Inc.). The planning dose distribution was then re-computed by replacing the true CT images with the SCT images under the same beam parameters in a quality assurance mode. The dose matrices have a resolution of $3\times3\times3$ mm^3 and covered the major regions of interest (ROI) including the parotids, brainstem, spinal cord, mandible, temporomandibular joint, and temporal lobes. The resulting dose matrices were compared using the global 3-dimensional γ analysis under the criteria of 2%/2 mm and 3%/3 mm (dose discrepancy/distance agreement) with surpassing areas with point doses less than 10% of the maximum dose. Within the ROIs, the mean of the absolute point dose discrepancies was calculated. Dose-volume histogram (DVH) parameters calculated from the true CT images and SCT images were compared.

Statistical analysis

All statistical analyses were performed with SPSS software (Version 20, SPSS Inc, Chicago, IL). Wilcoxon signed-rank tests were used to compare the performance of the cGAN and cycleGAN models. Specifically, the MAEs between the SCT images and the true CT images and the γ passing rates were tested. A p-value of less than 0.05 was considered statistically significant.

Results

The cGAN and cycleGAN were trained in 24.8–99.2 h and 51.3–205.2 h, respectively. Once the models have been trained, it took 2.19–5.47 s to generate an SCT volume image from a new MR volume image. With all training samples, the mean ± standard deviation (SD) of MAE between the SCT and true CT images within the

Table 1
HU discrepancies between the true CT image and the synthetic CT image generated from the models trained with all samples.

Mean ± SD (range) HU	ME		MAE	
	cGAN	cycleGAN	cGAN	cycleGAN
Body	18.37 ± 16.36 (−12.35, 55.34)	6.69 ± 19.41 (−46.62, 31.1)	69.67 ± 9.27 (49.48, 89.41)	100.62 ± 7.69 (82.96, 113.69)
Air (HU < −100)	23.9 ± 70.21 (−106, 188.68)	−78.66 ± 70.17 (−192.32, 64.25)	170.62 ± 36.38 (130.52, 256.7)	201.89 ± 34.11 (158.07, 292.43)
Soft tissue (150 > HU $\geq$ −100)	3.09 ± 11.45 (−27.13, 21.61)	−9.02 ± 12.22 (−42.41, 5.1)	38.09 ± 8.25 (23.11, 52.65)	51.94 ± 7.07 (38.6, 66.11)
Bone (HU $\geq$ 150)	95.66 ± 46.05 (−1.52, 180.07)	146.52 ± 56.59 (−51.66, 226.75)	203.71 ± 28.22 (158.59, 262.6)	288.17 ± 17.22 (261.77, 345.65)

Abbreviations: cGAN = conditional generative adversarial network; cycleGAN = cycle-consistent generative adversarial network; HU = Hounsfield unit; MAE = mean absolute error; ME = mean error; SD = standard deviation.

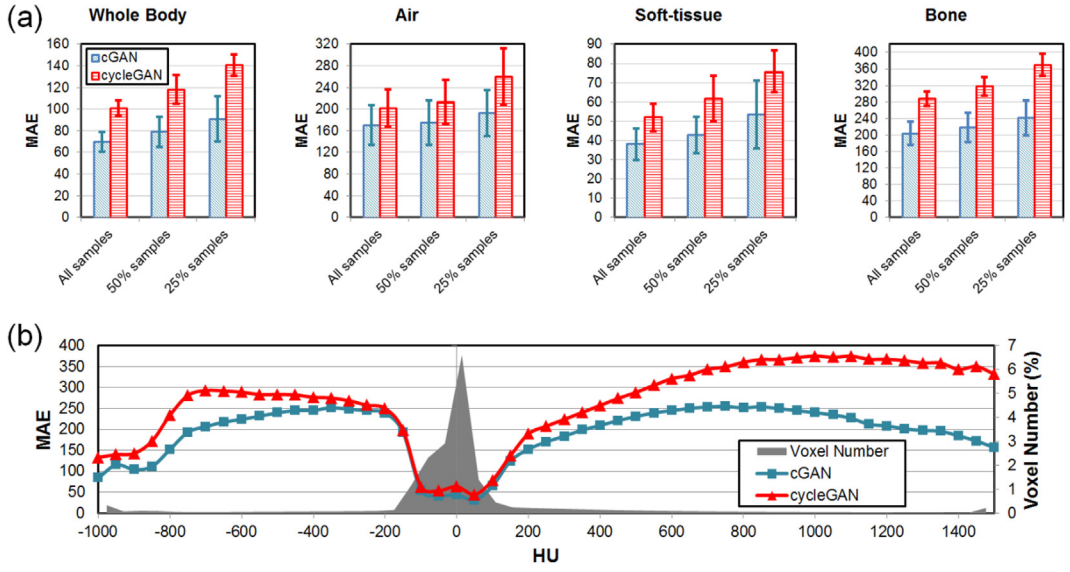

Fig. 2. (a) From left to right: the MAE within the body, air (HU < 200 HU), soft tissue (150 $\geq$ HU > −200), and bone (HU $\geq$ 150 HU) *versus* different numbers of training samples; (b) the MAE (left axis) calculated in each HU bin for the cGAN and cycleGAN trained with all samples; the HU histogram (right axis) of all voxels within the body. Abbreviation: cGAN: conditional generative adversarial network, cycleGAN: cycle-consistent generative adversarial network, MAE = mean absolute error.

body were (69.67 ± 9.27) HU and (100.62 ± 7.69) HU for the cGAN and cycleGAN, respectively. The details of the corresponding ME and MAE in different tissues are listed in [Table 1]. With the training samples reduced to 50% and 25%, the MAEs within the body increased to (79.05 ± 13.88) HU and (91.02 ± 21) HU for the cGAN, whereas it increased to (118.16 ± 13.53) HU and (140.84 ± 9.77) HU, respectively, for the cycleGAN. The ME and MAE in different tissues are detailed in [Supplement Table A and B]. Fig. 2(a) shows the MAEs *versus* different numbers of training samples. Fig. 2(b) shows the MAE calculated in each HU bin for the two models trained with all samples. The cycle GAN resulted in a much larger MAE for the high-density bone (>500 HU) and air (<200 HU). Fig. 3 shows the SCT images generated from the models trained with all samples and showed average performance (cGAN MAE = 68.44 HU; cycleGAN MAE = 94.01 HU) and worst performance on HU estimation (cGAN MAE = 89.41 HU; cycleGAN MAE = 108.05 HU).

Within the body, the (mean ± SD) of 2%/2 mm and 3%/3 mm γ pass rates for the cGAN-generated SCT images were (98.68 ± 0.94)% and (99.64 ± 0.33)%, whereas they were (98.52 ± 1.13)% and (99.59 ± 0.38)% for the cycleGAN-generated SCT. The 2%/2 mm and 3%/3 mm γ pass rates were over 96.9% and 98.78% for the cGAN-generated SCT images and were over 95.61% and 98.55% for the cycleGAN-generated SCT images. There was no significant difference in the γ pass rates between the two models ($p > 0.05$). Fig. 4(a) and (b) shows the true CT images, cGAN-generated SCT images, and calculated dose distributions for a representative case of worse dosimetric performance. The DVHs were also displayed. Fig. 4(c)and (d) shows the worst dosimetric performance in the cycleGAN-generated SCT images. The corresponding DVHs are shown as well. Fig. 5 shows the mean of absolute point dose discrepancies (% of the prescription) within ROIs and the absolute discrepancies of the DVH parameters. Overall, the cGAN achieved a relatively smaller dosimetric discrepancy.

Discussion

Generating SCT images from conventional MR image remains challenging in air-bearing bone structures. In this study, the SCT image generated by the cGAN achieved lower MAE than did those generated by the cycleGAN, especially for high-density bony tissues. The results demonstrated that although cycleGAN can generate real-looking CT images, regions with low MR signals, such as bone/air interfaces, can cause uncertainties [Supplementary Figure A]. Our results are consistent with those of a recently published study on SCT image generation using mDixon MR image for head and neck cancer patients [36]. The study demonstrated that the cycleGAN achieved worse performance than the cGAN. However, another study demonstrated that cycleGAN generated better SCT images than did cGAN [31]. The discrepancies could be because the latter study used rigid registration for creating the paired training samples and did not account for the tissue/organ deformations between the CT and MR images. We observed that such deformations were not negligible in head and neck regions even though the MR and CT images were acquired with an interval time of less than 2 h [Supplementary Figure B]. Additionally, the geometry distortion on the MR image, especially the subject-induced suscepti-

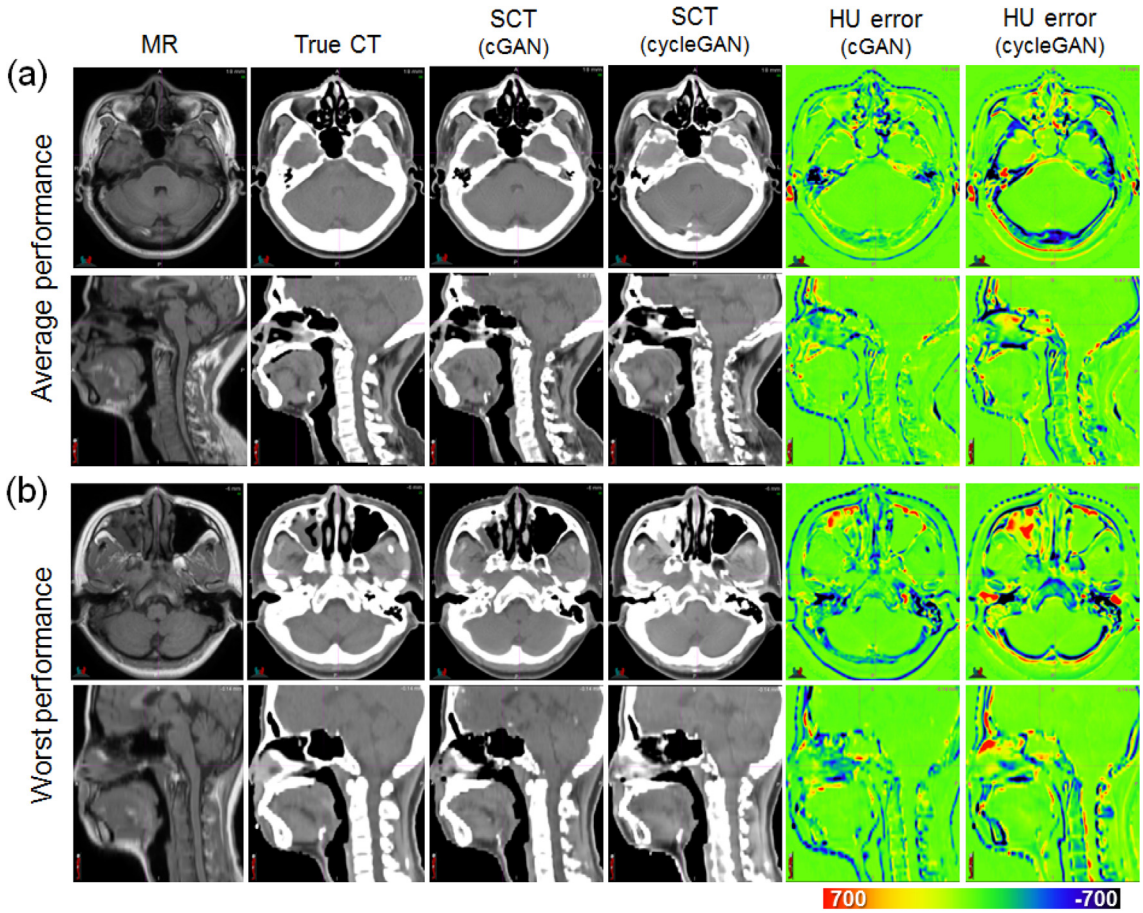

Fig. 3. (a) Case with average HU accuracy in the generated SCT images (cGAN MAE = 68.44 HU; cycleGAN MAE = 94.01 HU). Case with the worst performance (cGAN MAE = 89.41 HU; cycleGAN MAE = 108.95 HU). (b) The SCT images were from the models trained with all samples. HU error = SCT – true CT. *Abbreviations*: cGAN = conditional generative adversarial network; cycleGAN = cycle-consistent generative adversarial network; MAE = mean absolute error; SCT = synthetic CT.

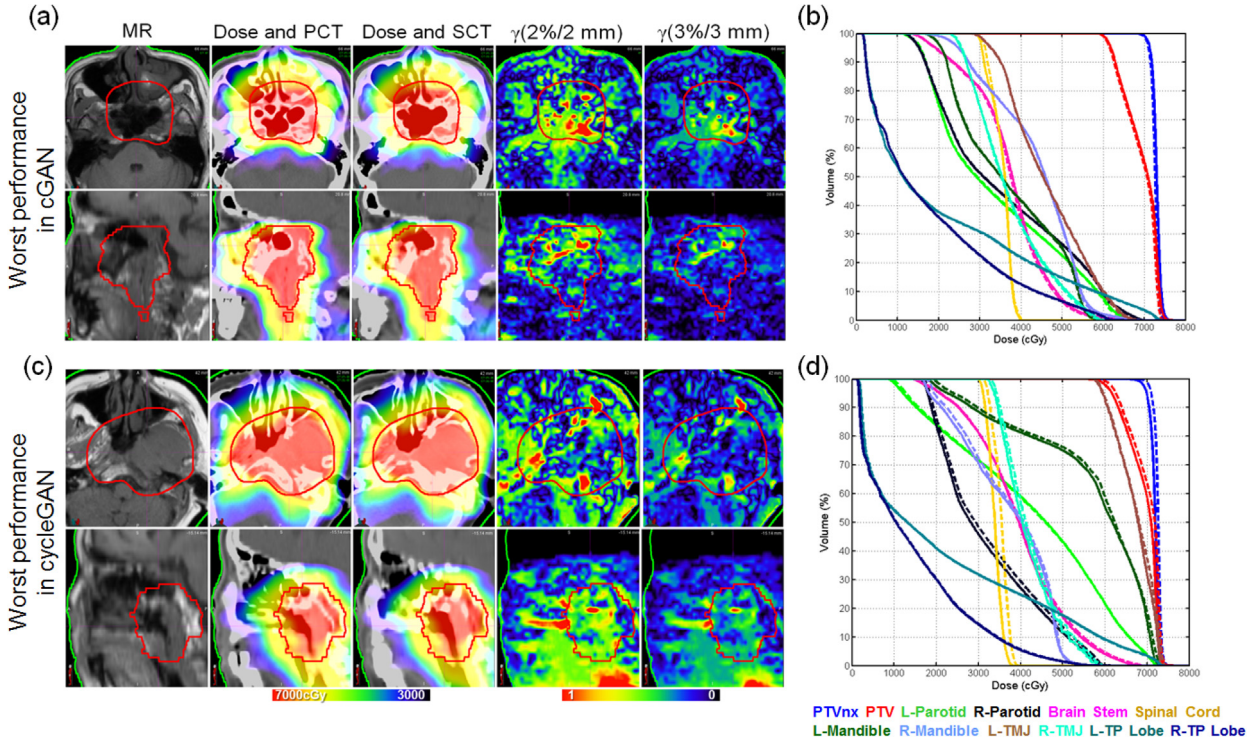

Fig. 4. Worst case on dosimetric accuracy in the cGAN-generated SCT images (3 mm/3% γ passing rate = 98.78%) (a) and its DVH (b). Worst case in cycleGAN-generated SCT images (3 mm/3% γ passing rate = 98.55%) (c) and its DVH (d). *Abbreviations*: cGAN = conditional generative adversarial network; cycleGAN = cycle-consistent generative adversarial network; DVH = dose-volume histogram; SCT = synthetic CT; PTVnx: planning target volume of primary disease, PTV: planning target volume (the red contours in (a) and (c)); TMJ: temporomandibular joint, TP lobe: temporal lobe. L: left, R: right.

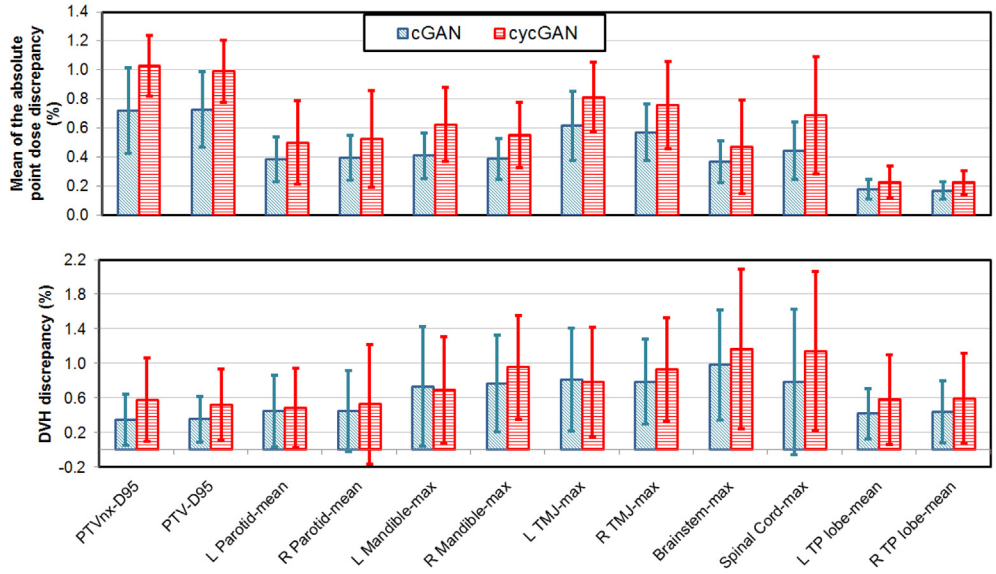

Fig. 5. Mean of the absolute point dose discrepancy (in % of the prescription) (a) and the absolute discrepancy for the DVH parameters (in %) (b). *Abbreviations*: cGAN = conditional generative adversarial network; cycleGAN = cycle-consistent generative adversarial network; DVH = dose-volume histogram; TMJ: temporomandibular joint; TP: temporal lobe; PTVnx: planning target volume of primary disease; PTV: planning target volume; L: left, R: right.

bility distortion near tissues/air interfaces [57], may cause a local mismatch between MR and CT images. In such situations, rigid registration may not be sufficient to create perfectly matched MR-CT image pairs for training deep CNN or cGAN.

The MAEs increased with decreasing training samples for both cGAN and cycleGAN. With 25% of training samples (34 MR-CT pairs), the cGAN achieved a MAE = (91.02 ± 21) HU within the body. A previous study used 23 MR-CT pairs from NPC patients to train a U-net [24] and reported a MAE = (131 ± 24) HU within the body. These results indicated the feasibility of reducing the training samples (e.g., to 30 MR-CT pairs) for cGAN or U-net while achieving a reasonable SCT image estimation accuracy. With the maximum training samples (135 pairs), the cGAN achieved the lowest MAE = (69.97 ± 9.27) HU within the body, which was even

lower than that in previous studies [47,49] that used multiple MR sequence images to generate SCT image for NPC patients. These results demonstrated that the performance of cGAN could be further improved if more pairs were available. However, the results from other studies may not be directly comparable because the imaging protocol and scanner parameters were different.

The SCT images generated by both cGAN and cycleGAN achieved high dosimetric accuracy, with the pass rates much higher than the typical criteria for treatment delivery quality assurance. These findings support the feasibility of using single conventional MR sequence imaging for NPC IMRT planning. Further, the cGAN achieved higher accuracy in predicting the CT number for high-density bone than did cycleGAN. However, the dosimetric improvement was small. The reason could be those high-density voxels accounted for only a small portion of all the image voxels and had no profound effect on the overall dosimetric accuracy. However, the dosimetric uncertainty caused by local tissue heterogeneity was not negligible for both cGAN and cycleGAN. The discrepancy in the intensity of the isolated image voxels within the nasopharynx as the target organ can result in a local dose grid failing to pass 2%/2 mm γ criteria. These findings suggested that further improvement is possible by including specific MR sequence images, such as mDixon [36,58] to more accurately differentiate between the bone and air on the MR image.

In conclusion, both cGAN and cycleGAN could swiftly generate accurate SCT volume images from T1-weighted MR images and achieve clinically acceptable dosimetric accuracy for NPC IMRT planning. cGAN was preferable over cycleGAN if high-quality MR-CT image pairs were available. Further research is needed to evaluate the impact of MRI geometric distortion on treatment planning before implementing MRI-only NPC radiotherapy in the clinic.

Conflict of interest

None.

Acknowledgements

This work was jointly supported by the National Key R&D Program of China (2017YFC0113200), Science and Technology Program of Guangdong Province, China (2015B020214002), Science and Technology Program of Guangzhou, China (201607010199).

Ethics statement

Our study was reviewed and approved by the IRB committee of Sun Yat-sen University Cancer Center, with the approval number of CZR2020-042. As this study is only a retrospective analysis, informed consent exemption was approved (show in the attached IRB document).

Data statement

The datasets are backed up on the Research Data Deposit (RDD Number: RDDB2019000772, https://www.researchdata.org.cn) and are available upon reasonable request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.radonc.2020.06.049.

References

[1] Debois M, Oyen R, Maes F, Verswijvel G, Gatti G, Bosmans H, et al. The contribution of magnetic resonance imaging to the three-dimensional treatment planning of localized prostate cancer. Int J Radiat Oncol Biol Phys 1999;45:857–65. doi:S0360-3016(99)00288-6 [pii].
[2] Tenhunen M, Korhonen J, Kapanen M, Seppälä T, Koivula L, Collan J, et al. MRI-only based radiation therapy of prostate cancer: workflow and early clinical experience. Acta Oncol 2018;57:902–7. https://doi.org/10.1080/0284186X.2018.1445284.
[3] Kapanen M, Collan J, Beule A, Seppälä T, Saarilahti K, Tenhunen M. Commissioning of MRI-only based treatment planning procedure for external beam radiotherapy of prostate. Magn Reson Med 2013;70:127–35. https://doi.org/10.1002/mrm.24459.
[4] Huynh T, Gao Y, Kang J, Wang L, Zhang P, Lian J, et al. Estimating CT image from MRI data using structured random forest and auto-context model. IEEE Trans Med Imaging 2016;35:174–83. https://doi.org/10.1109/TMI.2015.2461533.
[5] Cao X, Yang J, Gao Y, Guo Y, Wu G, Shen D. Dual-core steered non-rigid registration for multi-modal images via bi-directional image synthesis. Med Image Anal 2017;41:18–31. https://doi.org/10.1016/j.media.2017.05.004.
[6] Dowling J A., Lambert J, Parker J, Salvado O, Fripp J, Capp A, et al. An atlas-based electron density mapping method for magnetic resonance imaging (MRI)-alone treatment planning and adaptive MRI-based prostate radiation therapy. Int J Radiat Oncol Biol Phys 2012;83:e5–11. doi:10.1016/j.ijrobp.2011.11.056.
[7] Uh J, Merchant TE, Li Y, Li X, Hua C. MRI-based treatment planning with pseudo CT generated through atlas registration. Med Phys 2014;41:. https://doi.org/10.1118/1.4873315051711.
[8] Gudur MSR, Hara W, Le Q-T, Wang L, Xing L, Li R. A unifying probabilistic Bayesian approach to derive electron density from MRI for radiation therapy treatment planning. Phys Med Biol 2014;59:6595–606. https://doi.org/10.1088/0031-9155/59/21/6595.
[9] Dowling JA, Sun J, Pichler P, Rivest-Hénault D, Ghose S, Richardson H, et al. Automatic substitute computed tomography generation and contouring for magnetic resonance imaging (MRI)-alone external beam radiation therapy from standard MRI sequences. Int J Radiat Oncol Biol Phys 2015;93:1144–53. https://doi.org/10.1016/j.ijrobp.2015.08.045.
[10] Andreasen D, Van Leemput K, Hansen RH, Andersen J a. L, Edmund JM. Patch-based generation of a pseudo CT from conventional MRI sequences for MRI-only radiotherapy of the brain. Med Phys 2015;42:1596–605. doi:10.1118/1.4914158
[11] Siversson C, Nordström F, Nilsson T, Nyholm T, Jonsson J, Gunnlaugsson A, et al. Technical Note: MRI only prostate radiotherapy planning using the statistical decomposition algorithm. Med Phys 2015;42:6090–7. https://doi.org/10.1118/1.4931417.
[12] Chen S, Quan H, Qin A, Yee S, Yan D. MR image-based synthetic CT for IMRT prostate treatment planning and CBCT image-guided localization. J Appl Clin Med Phys 2016;17:236–45.
[13] Demol B, Boydev C, Korhonen J, Reynaert N. Dosimetric characterization of MRI-only treatment planning for brain tumors in atlas-based pseudo-CT images generated from standard T 1-weighted MR images. Med Phys 2016;43:6557–68. https://doi.org/10.1118/1.4967480.
[14] Largent A, Barateau A, Nunes JC, Lafond C, Greer PB, Dowling JA, et al. Pseudo-CT generation for MRI-only radiation therapy treatment planning: comparison among patch-based, atlas-based, and bulk density methods. Int J Radiat Oncol Biol Phys 2019;103:479–90. https://doi.org/10.1016/j.ijrobp.2018.10.002.
[15] Ravishankar H, Sudhakar P, Venkataramani R, Thiruvenkadam S, Annangi P, Babu N, et al. Estimating CT image from MRI data using 3D fully convolutional networks. LABELS 2016;1:170–8. https://doi.org/10.1007/978-3-319-46976-8.
[16] Han L, Dong H, McClelland JR, Han L, Hawkes DJ, Barratt DC. A hybrid patient-specific biomechanical model based image registration method for the motion estimation of lungs. Med Image Anal 2017;39:87–100.
[17] Spadea MF, Pileggi G, Zaffino P, Salome P, Catana C, Izquierdo-Garcia D, et al. Deep convolution neural network (DCNN) multiplane approach to synthetic CT generation from MR IMAGES—application in brain proton therapy. Int J Radiat Oncol Biol Phys 2019;105:495–503. https://doi.org/10.1016/j.ijrobp.2019.06.2535.
[18] Chen S, Qin A, Zhou D, Yan D. Technical Note: U-net-generated synthetic CT images for magnetic resonance imaging-only prostate intensity-modulated radiation therapy treatment planning. Med Phys 2018;45:5659–65. https://doi.org/10.1002/mp.13247.
[19] Xiang L, Wang Q, Nie D, Zhang L, Jin X, Qiao Y, et al. Deep embedding convolutional neural network for synthesizing CT image from T1-weighted MR image. Med Image Anal 2018;47:31–44. https://doi.org/10.1016/j.media.2018.03.011.
[20] Gupta D, Kim M, Vineberg KA, Balter JM. Generation of synthetic CT images from MRI for treatment planning and patient positioning using a 3-channel U-Net trained on sagittal images. Front Oncol 2019;9:1–8. https://doi.org/10.3389/fonc.2019.00964.
[21] Dinkla AM, Florkow MC, Maspero M, Savenije MHF, Zijlstra F, Doornaert PAH, et al. Dosimetric evaluation of synthetic CT for head and neck radiotherapy generated by a patch-based three-dimensional convolutional neural network. Med Phys 2019;46. https://doi.org/10.1002/mp.13663.
[22] Arabi H, Dowling JA, Burgos N, Han X, Greer PB, Koutsouvelis N, et al. Comparative study of algorithms for synthetic CT generation from MRI: consequences for MRI-guided radiation planning in the pelvic region. Med Phys 2018;45:5218–33. https://doi.org/10.1002/mp.13187.

[23] Dinkla AM, Wolterink JM, Maspero M, Savenije MHF, Verhoeff JJC, Seravalli E, et al. MR-only brain radiotherapy: dosimetric evaluation of synthetic CTs generated by a dilated convolutional neural network. Int J Radiat Oncol Biol Phys 2018;102:801–12. https://doi.org/10.1016/j.ijrobp.2018.05.058.
[24] Wang Y, Liu C, Zhang X, Deng W. Synthetic CT generation based on T2 weighted MRI of nasopharyngeal carcinoma (NPC) using a deep convolutional neural network (DCNN). Front Oncol 2019;9. https://doi.org/10.3389/fonc.2019.01333.
[25] Neppl S, Landry G, Kurz C, Hansen DC, Hoyle B, Stöcklein S, et al. Evaluation of proton and photon dose distributions recalculated on 2D and 3D Unet-generated pseudoCTs from T1-weighted MR head scans. Acta Oncol 2019;58:1429–34. https://doi.org/10.1080/0284186X.2019.1630754.
[26] Nie D, Trullo R, Lian J, Petitjean C, Ruan S, Wang Q, et al. Medical image synthesis with deep convolutional adversarial networks. IEEE Trans Biomed Eng 2017;9294:417–25. https://doi.org/10.1007/978-3-319-66179-7_48.
[27] Maspero M, Savenije MHF, Dinkla AM, Seevinck PR, Intven MPW, Juergenliemk-Schulz IM, et al. Dose evaluation of fast synthetic-CT generation using a generative adversarial network for general pelvis MR-only radiotherapy. Phys Med Biol 2018;63. https://doi.org/10.1088/1361-6560/aada6d.
[28] Emami H, Dong M. Generating synthetic CTs from magnetic resonance images using generative adversarial networks. Med Phys 2018;45:3627–36.
[29] Olberg S, Zhang H, Kennedy WR, Chun J, Rodriguez V, Zoberi I, et al. Synthetic CT reconstruction using a deep spatial pyramid convolutional framework for MR-only breast radiotherapy. Med Phys 2019. https://doi.org/10.1002/mp.13716.
[30] Kazemifar S, McGuire S, Timmerman R, Wardak Z, Nguyen D, Park Y, et al. MRI-only brain radiotherapy: assessing the dosimetric accuracy of synthetic CT images generated using a deep learning approach. Radiother Oncol 2019;136:56–63. https://doi.org/10.1016/j.radonc.2019.03.026.
[31] Wolterink JM, Dinkla AM, Savenije MHF, Seevinck PR, van den Berg CAT, Isgum I. Deep MR to CT synthesis using unpaired data. MICCAI 2017:1–10.
[32] Xiang L, Li Y, Lin W, Wang Q. Unpaired Deep Cross-Modality Synthesis with Fast Training. DLMIA, vol. 10553, Springer International Publishing; 2018. doi:10.1007/978-3-319-67558-9.
[33] Hiasa Y, Otake Y, Takao M, Matsuoka T, Takashima K, Prince JL, et al. Cross-modality image synthesis from unpaired data using CycleGAN Effects of gradient consistency loss and training data size. ArXiv Prepr ArXiv180306629 2018:1–8.
[34] Yang H, Sun J, Carass A, Zhao C, Lee J, Xu Z, et al. Unpaired brain MR-to-CT synthesis using a structure-constrained CycleGAN. MICCAI 2018:4–7.
[35] Wu H, Jiang X, Jia F. UC-GAN for MR to CT image synthesis. In: Nguyen D, Xing L, Jiang S, editors. MICCAI. Cham: Springer International Publishing; 2019. p. 146–53.
[36] Klages P, Bensilmane I, Riyahi S, Jiang J, Hunt M, Deasy JO, et al. Patch-based generative adversarial neural network models for head and neck MR-Only planning. Med Phys 2019. https://doi.org/10.1017/CBO9781107415324.004.
[37] Lei Y, Harms J, Wang T, Liu Y, Shu HK, Jani AB, et al. MRI-only based synthetic CT generation using dense cycle consistent generative adversarial networks. Med Phys 2019;46:3565–81. https://doi.org/10.1002/mp.13617.
[38] Han X. MR-based synthetic CT generation using a deep convolutional neural network method. Med Phys 2017;44:1408–19. https://doi.org/10.1002/mp.12155.
[39] Pathak D, Krahenbuhl P, Donahue J, Darrell T, Efros AA. Context encoders: feature learning by inpainting. CVPR, vol. 2016- Decem, 2016, p. 2536–44. doi:10.1109/CVPR.2016.278.
[40] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative Adversarial Nets. NIPS, 2014, p. 2672–80. doi:10.1017/CBO9781139058452.
[41] Isola P, Zhu J-Y, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. CVPR, 2017. doi:arXiv:1611.07004
[42] Zhu J, Park T, Isola P, Efros AA. Unpaired image-to-mage translation using cycle-consistent adversarial networks. ICCV 2017:2223–32. https://doi.org/10.1109/ICCV.2017.244.
[43] Largent A, Barateau A, Nunes JC, Mylona E, Castelli J, Lafond C, et al. Comparison of deep learning-based and patch-based methods for pseudo-CT generation in MRI-based prostate dose planning. Int J Radiat Oncol Biol Phys 2019;105:1137–50. https://doi.org/10.1016/j.ijrobp.2019.08.049.
[44] Liu Y, Lei Y, Wang T, Kayode O, Tian S, Liu T, et al. MRI-based treatment planning for proton radiotherapy: dosimetric validation of a deep learning-based liver synthetic CT generation method. Phys Med Biol 2019;64. https://doi.org/10.1259/bjr.20190067.
[45] Liu Y, Lei Y, Wang T, Kayode O, Tian S, Liu T, et al. MRI-based treatment planning for liver stereotactic body radiotherapy: validation of a deep learning-based synthetic CT generation method. Br J Radiol 2019;92:20190067.
[46] Fu J, Singhrao K, Cao M, Yu V, Santhanam AP, Yang Y, et al. Generation of abdominal synthetic CTs from 0.35 T MR images using generative adversarial networks for MR-only liver radiotherapy. Biomed Phys Eng Express 2020;6:15033.
[47] Tie X, Lam S, Zhang Y, Lee K, Au K, Cai J. Pseudo-CT generation from multi-parametric MRI using a novel multi-channel multi-path conditional generative adversarial network for nasopharyngeal carcinoma patients. Med Phys 2020. https://doi.org/10.1002/mp.14062.
[48] Klages P, Benslimane I, Riyahi S, Jiang J, Hunt M, Deasy JO, et al. Patch-based generative adversarial neural network models for head and neck MR-only planning. Med Phys 2020;47:626–42. https://doi.org/10.1002/mp.13927.
[49] Qi M, Li Y, Wu A, Jia Q, Li B, Sun W, et al. Multi-sequence MR image-based synthetic CT generation using a generative adversarial network for head and neck MRI-only radiotherapy. Med Phys 2020. https://doi.org/10.1002/mp.14075.
[50] Han X, Hibbard LS, Willcut V. An efficient inverse-consistent diffeomorphic image registration method for prostate adaptive radiotherapy. Prostate Cancer Imaging. Springer; 2010. p. 34–41.
[51] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. MICCAI 2015. https://doi.org/10.1007/978-3-319-24574-4_28.
[52] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. ICML 2015. https://doi.org/10.1007/s13398-014-0173-7.2.
[53] Ulyanov D, Vedaldi A, Lempitsky V. Instance normalization: the missing ingredient for fast stylization. ArXiv:160708022v3 2017.
[54] Maas AL, Hannun AY, Ng AY. Rectifier nonlinearities improve neural network acoustic models. ICML 2013;28:6.
[55] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. CVPR 2016. https://doi.org/10.1109/CVPR.2016.90.
[56] Mao X, Li Q, Xie H, Lau RYK, Wang Z, Smolley SP. Multi-class Generative Adversarial Networks with the L2 Loss Function. ArXiv Prepr ArXiv161104076 5 2016. doi:10.1109/ICCV.2017.304.
[57] Glide-hurst C, Nejad-davarani S, Weiss S, Zheng W, Chetty IJ, Renisch S. Per-organ assessment of subject-induced susceptibility distortion for MR-only male pelvis treatment planning. Radiat Oncol 2018:1–9. https://doi.org/10.1186/s13014-018-1090-2.
[58] Farjam R, Tyagi N, Veeraraghavan H, Apte A, Zakian K, Hunt MA, et al. Multiatlas approach with local registration goodness weighting for MRI-based electron density mapping of head and neck anatomy. Med Phys 2017:6–12. https://doi.org/10.1002/mp.12303.