PAPER

Generating synthesized computed tomography (CT) from cone-beam computed tomography (CBCT) using CycleGAN for adaptive radiation therapy

To cite this article: Xiao Liang *et al* 2019 *Phys. Med. Biol.* **64** 125002

View the article online for updates and enhancements.

Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

CrossMark

PAPER

Generating synthesized computed tomography (CT) from cone-beam computed tomography (CBCT) using CycleGAN for adaptive radiation therapy

Xiao Liang[1,2], Liyuan Chen[1,2], Dan Nguyen[1], Zhiguo Zhou[1], Xuejun Gu[1], Ming Yang[1], Jing Wang[1] and Steve Jiang[1]

[1] Department of Radiation Oncology, Medical Artificial Intelligence and Automation Laboratory, University of Texas Southwestern Medical Center, Dallas, TX, United States of America
[2] Co-first authors.

E-mail: Steve.Jiang@UTSouthwestern.edu and Jing.Wang@UTSouthwestern.edu

Abstract

Throughout the course of delivering a radiation therapy treatment, which may take several weeks, a patient's anatomy may change drastically, and adaptive radiation therapy (ART) may be needed. Cone-beam computed tomography (CBCT), which is often available during the treatment process, can be used for both patient positioning and ART re-planning. However, due to the prominent amount of noise, artifacts, and inaccurate Hounsfield unit (HU) values, the dose calculation based on CBCT images could be inaccurate for treatment planning. One way to solve this problem is to convert CBCT images to more accurate synthesized CT (sCT) images. In this work, we have developed a cycle-consistent generative adversarial network framework (CycleGAN) to synthesize CT images from CBCT images. This model is capable of image-to-image translation using unpaired CT and CBCT images in an unsupervised learning setting. The sCT images generated from CBCT through this CycleGAN model are visually and quantitatively similar to real CT images with decreased mean absolute error (MAE) from 69.29 HU to 29.85 HU for head-and-neck (H&N) cancer patients. The dose distributions calculated on the sCT by CycleGAN demonstrated a higher accuracy than those on CBCT in a 3D gamma index analysis with increased gamma index pass rate from 86.92% to 96.26% under 1 mm/1% criteria, when using the deformed planning CT image (dpCT) as the reference. We also compared the CycleGAN model with other unsupervised learning methods, including deep convolutional generative adversarial networks (DCGAN) and progressive growing of GANs (PGGAN), and demonstrated that CycleGAN outperformed the other two models. A phantom study has been conducted to compare sCT with dpCT, and the increase of structural similarity index from 0.91 to 0.93 shows that CycleGAN performed better than DIR in terms of preserving anatomical accuracy.

1. Introduction

Inter-factional anatomical change occurs to many cancer patients under radiotherapy and in many tumor sites such as head-and-neck (H&N) cancer, often due to tumor shrinkage or weight loss during treatment course. Relying solely on computed tomography (CT) images acquired before treatment may increase the risk of tumor underdose and organs-at-risk (OARs) overdose. Several studies have shown compromised tumor coverage caused by anatomical changes during treatment courses. One pilot study on H&N anatomical changes showed that gross tumor volume (GTV) decreased at a median rate of 1.8% per treatment day, with a median total relative loss of 69.5% on the last day of treatment and a median mass displacement of 3.3 mm. Also, the parotid gland showed a median reduction in volume of 0.19 cm^3 d^{-1} and a median shift of 3.1 mm (Barker *et al* 2004). Another study showed that H&N patients with significant anatomical changes like neck diameter reduction greater than 10% or weight loss more than 5% would be more likely to suffer from deleterious side effects such as xerostomia

(You *et al* 2012). A retrospective study that included 13 H&N patients showed that the doses to 95% (D_{95}) of the planning target volume (PTV) were reduced in 92% of patients. Also, the maximum dose (D_{max}) to the spinal cord increased in all patients, and the D_{max} to the brainstem increased in 85% of patients because of anatomical changes (Hansen *et al* 2006).

Findings from clinical trials have indicated that adaptive radiation therapy (ART) improves the dosimetric quality of radiation therapy plans in H&N cancer patients. A clinical trial evaluating ART benefits on H&N patients acquired daily in-room CT-on-rails or cone-beam computed tomography (CBCT) images to track anatomical changes and then used deformable image registration (DIR) to align contours on new images for replanning, if needed (Schwartz *et al* 2013). In this study, doses on contralateral and ipsilateral parotid were reduced on all 22 patients. However CT-on-rails are not available in most clinics. In contrast, CBCT images are frequently taken during the treatment course for patient alignment. Thus, using CBCT in ART is more practical and efficient than CT-on-rails. Unlike CT, scatter contribution expressed as the scatter-to-primary ratio (SPR) in CBCT can be as high as 3 for H&N patients, compared with 0.2 in CT (Miracle and Mukherji 2009). Hounsfield unit (HU) values in CBCT are also different from those in CT. All of these factors will affect segmentation and dose calculation accuracy on CBCT. Therefore, CBCT correction needs to be made before use in ART.

Popular CBCT correction methods for ART include the analytical modeling methods, Monte Carlo (MC) simulation, CT-prior-based methods, histogram matching, and learning-based methods. Analytical modeling methods try to approximate scatter distribution in projection data by assuming that the scatter signal is a convolution function of primary signal and scatter kernel (Naimuddin *et al* 1987). The MC simulation-based methods (Xu *et al* 2015) are more robust than the analytical method, because they simulate photon transport in a flexible yet rigorous way. The prior-CT-based methods exploit prior information obtained from DIR of CT to CBCT (Zöllner *et al* 2017). CBCT HU values can also be converted to CT HU values by matching the histograms of CBCT with the histograms of CT for each slice via linear scaling (Abe *et al* 2017). A regression forest model was trained to correct CBCT images through patient-specific anatomical features extracted from aligned CT and CBCT images (Lei *et al* 2018). These methods have achieved limited success, and there is still room for improvement. The last three methods all need paired datasets, meaning that CBCT as input and CT as label must have the same anatomy. However, even CBCT and CT images acquired on the same day from the same patient would still have slightly anatomical differences, affecting accuracy of simple mapping from CBCT to CT. The most common practice is to deform the planning CT (pCT) though DIR to CBCT anatomy and then to use the deformed planning CT (dpCT) for dose calculation. Thus, dpCT keeps the same HU accuracy as CT. However, the anatomical accuracy of dpCT could be an issue, and DIR could yield incorrect contours, due to more pronounced anatomical changes and reduced soft-tissue contrast (Kurz *et al* 2016). To overcome the problems demonstrated previously, we would like to generate synthesized CT (sCT) images that have CT's HU values and CBCT's anatomies directly from CBCT images without using paired CT and CBCT images for training.

Deep learning has become a general-purpose solution for image-to-image translation. Hand-engineered mapping functions that would traditionally require complicated formulations are no longer needed (Isola *et al* 2017). Therefore, deep learning is a new promising method for translating CBCT to CT images without knowing their mapping functions. One way to use deep learning methods for image-to-image translation is supervised training with paired images. Another way is unsupervised training with unpaired images, like generative adversarial networks (GANs) (Goodfellow *et al* 2014), CycleGAN (Zhu *et al* 2017), and Triangle-GAN (Gan *et al* 2017). CycleGAN was first applied to converting magnetic resonance imaging (MRI) to CT on H&N sites (Wolterink *et al* 2017) and then applied to converting MRI to CT on pelvic sites with more training data and incorporating gradient consistency loss (Hiasa *et al* 2018).

In this study, our goal is to synthesize CT images from CBCT images that have CT image quality while keeping CBCT anatomy. For supervised learning, it is difficult to acquire paired CT and CBCT images with exactly matching anatomy. Due to this issue, we applied CycleGAN, an unsupervised training method, to directly convert CBCT to CT-like images. The CycleGAN model was used to learn translation functions from a source domain CBCT to a target domain CT. However, in the absence of paired images, training with an inverse mapping from CT to CBCT was coupled at the same time and a cycle consistency loss was introduced to constrain the mapping. This is the first time an unpaired CBCT and CT dataset was used to train a model, which overcomes the limitation of lacking paired CBCT and CT datasets in reality. Unlike DIR, the generated sCT images should keep exactly the same anatomy as CBCT, while having the HU accuracy of pCT. We compared the performance of CycleGAN with two other unsupervised deep learning models: deep convolutional generative adversarial network (DCGAN) and progressive growing of GANs (PGGAN), and analyzed the HU accuracy of the generated images from these three deep learning methods by similarity measures and 3D gamma index analysis of calculated dose distributions. Finally, we conducted a phantom study to compare the performance of the CycleGAN model with the DIR method in terms of preserving anatomical accuracy.

Figure 1. CycleGAN architecture is used to convert CBCT images to sCT images.
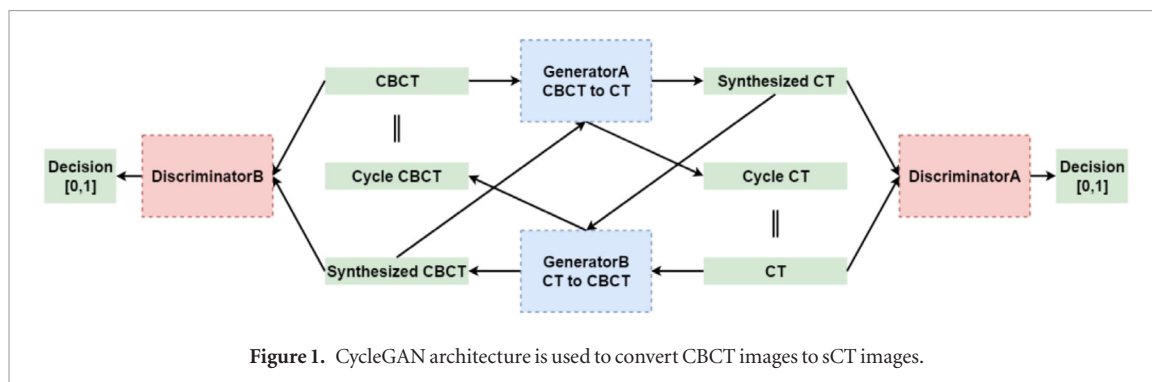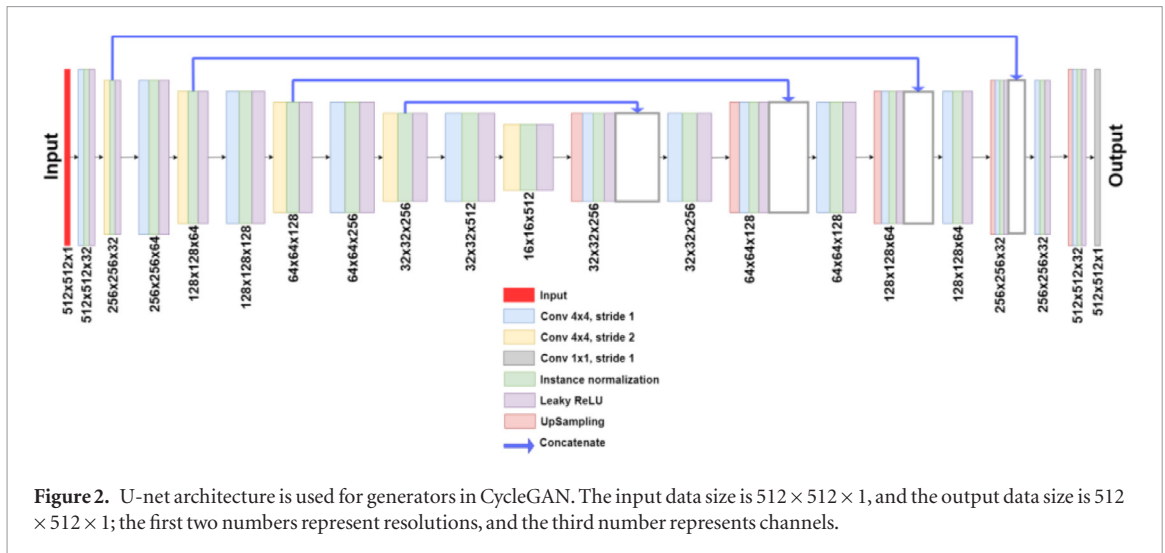
2. Materials and methods

2.1. Data

The pCT and CBCT images of H&N cancer patients were divided into 81 training, 9 validation, and 20 test patients. Patients were randomly picked from one clinical database, and the pCT and CBCT images for each patient were scanned on different days. CBCT volumes were acquired by Varian On-Board Imager with voxel spacing of $0.51 \times 0.51 \times 1.99\,\text{mm}^3$ and dimensions of $512 \times 512 \times 93$. The pCT volumes were acquired by a Philips CT scanner with multi-slice acquisition mode and with voxel spacing of $1.17 \times 1.17 \times 3.00\,\text{mm}^3$ and dimensions of $512 \times 512 \times 140$. Due to GPU memory limitations, a 2D model was trained and tested. Thus, 6480 and 720 2D image slices only from the axial plane were selected from 81 and 9 CT volumes for training and validation, respectively. The CT slices used for training were resampled to CBCT's pixel spacing and then cropped to 512×512 dimensions. All images in the training dataset were normalized to $(-1, 1)$ range. During the training stage, CBCT and CT slices were randomly shuffled at each epoch so that the correspondence to patients was removed. The validation dataset was used for hyperparameter tuning, and the test dataset was untouched during training. For testing, Elastix (Klein *et al* 2010), an open source image registration toolbox, was used to deform pCT to CBCT's anatomy to generate dpCT images as the ground truth for evaluating sCT HU value accuracy for the patient study. For the phantom study, real CT images were used as the ground truth to compare the performance of the CycleGAN model with the DIR method.

2.2. Overview of CycleGAN

The CycleGAN model structure is shown in figure 1. Like the other GAN models, CycleGAN includes generators and discriminators that compete against one another until they converge to an equilibrium point. Discriminators are used to distinguish between real and fake data, while generators are used to generate fake data with a distribution that eventually matches that of the real data, tricking the discriminators to classify the fake data as real. Two generators and two discriminators are included in CycleGAN. GeneratorA (G_A) generates sCT from CBCT, and GeneratorB (G_B) generates synthesized CBCT (sCBCT) from CT. DiscriminatorA (D_A) discriminates between CT and sCT, while DiscriminatorB (D_B) discriminates between CBCT and sCBCT. Two cycles are included in CycleGAN. In the first cycle, CBCT is the input into G_A and sCT is the output. Then, G_B takes sCT as input and generates cycle CBCT, which is supposed to be equal to CBCT. Meanwhile, D_A discriminates between CT and sCT by giving label 1 to CT and label 0 to sCT. In the second cycle, CT is the input into G_B and sCBCT is the output. Then, G_A takes sCBCT as input and generates cycle CT, which is supposed to be equal to CT. Meanwhile, D_B discriminates between CBCT and sCBCT by giving label 1 to CBCT and label 0 to sCBCT.

2.3. Generators and discriminators of CycleGAN

All models are trained with Adam optimization with a mini-batch size of 2. All weights are initialized from a random normal initializer with a mean of 0 and a standard deviation of 0.02. Instance normalization is used after convolution, and LeakyReLU with slope 0.2 is used after instance normalization in all models for fast convergence (Wang *et al* 2018). The learning rate is set to 0.0002. We set the momentum term β_1 to 0.5 to stabilize training (Radford *et al* 2015). We use a U-Net architecture for generators (figure 2) and 142×142 patchGAN for discriminators (figure 3). A 70×70 patchGAN was used in image-to-image translation for 256×256 image resolution (Isola *et al* 2017), because the full ImageGAN not only does not improve the visual quality of the results, but may even generate lower quality. We chose a 142×142 patchGAN as our discriminators for our 512×512 image resolution. 142×142 is the receptive field, which is the same size as the input region to the network, which contributes to layer activation. The last layer activation function in the generators is the Tanh activation function, and the last layer activation function in the discriminators is the linear activation function. The validation dataset is used to tune hyper-parameters.

Figure 2. U-net architecture is used for generators in CycleGAN. The input data size is $512 \times 512 \times 1$, and the output data size is $512 \times 512 \times 1$; the first two numbers represent resolutions, and the third number represents channels.

2.4. Loss function of CycleGAN

As shown in figure 1, CycleGAN includes two mapping functions: mapping from CBCT to CT and mapping from CT to CBCT. Additionally, two discriminators, D_A and D_B, are also included in CycleGAN. D_A aims to distinguish pCT from fake CT, and D_B aims to distinguish real CBCT from fake CBCT. Our objective includes three types of terms: adversarial loss for mapping the distribution of the generated images to the distribution of the target images; cycle consistency loss to prevent the two mappings from contradicting each other; and identity mapping loss to help preserve the gray level of the target images on generated images (Zhu *et al* 2017).

Adversarial loss is reflected in both the generators and the discriminators. D_A aims to classify sCT with label 0 from pCT with label 1, while G_A aims to generate sCT with a label approaching the pCT label. Therefore, D_A tries to minimize $\mathcal{L}_{GAN-D_A}$,

$$\mathcal{L}_{GAN-D_A} = \frac{1}{m}\sum\nolimits_{i=1}^{m} \frac{\left(1 - D_A\left(CT_i\right)\right)^2 + D_A(G_A\left(CBCT_i\right))^2}{2}, \tag{1}$$

while G_A tries to minimize $\mathcal{L}_{GAN-G_A}$,

$$\mathcal{L}_{GAN-G_A} = \frac{1}{m}\sum_{i=1}^{m} \left(1 - D_A\left(G_A\left(CBCT_i\right)\right)\right)^2. \tag{2}$$

Similarly, D_B classifies sCBCT with label 0 from real CBCT with label 1. G_B generates sCBCT with a label approaching the real CBCT label. Therefore, D_B tries to minimize $\mathcal{L}_{GAN-D_B}$,

$$\mathcal{L}_{GAN-DB} = \frac{1}{m}\sum\nolimits_{i=1}^{m} \frac{\left(1 - D_B\left(CBCT_i\right)\right)^2 + D_B(G_B\left(CT_i\right))^2}{2}, \tag{3}$$

while G_B tries to minimize $\mathcal{L}_{GAN-G_B}$,

$$\mathcal{L}_{GAN-G_B} = \frac{1}{m}\sum_{i=1}^{m} \left(1 - D_B\left(G_B\left(CT_i\right)\right)\right)^2. \tag{4}$$
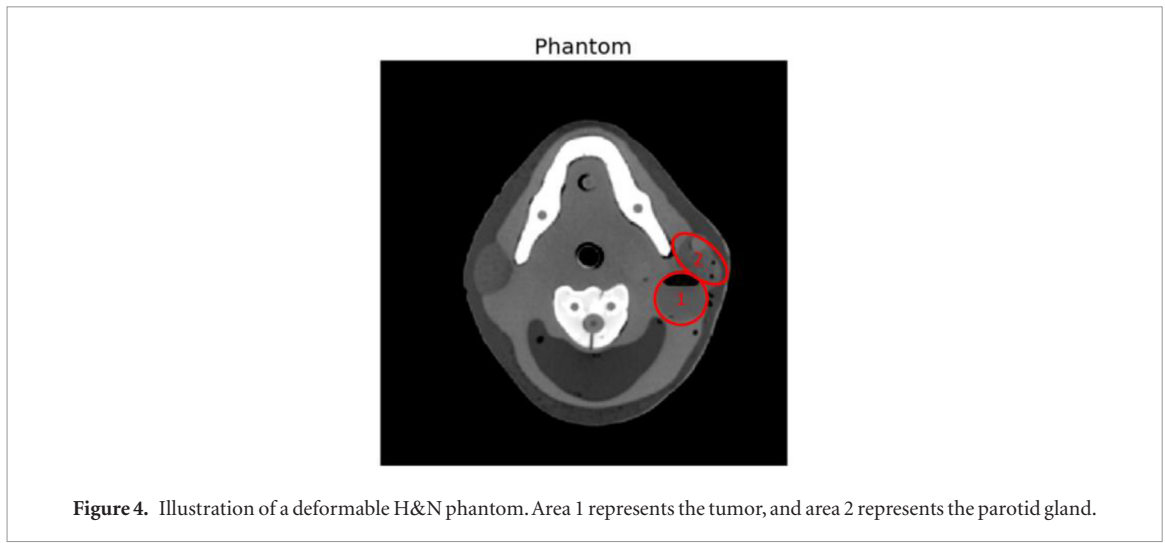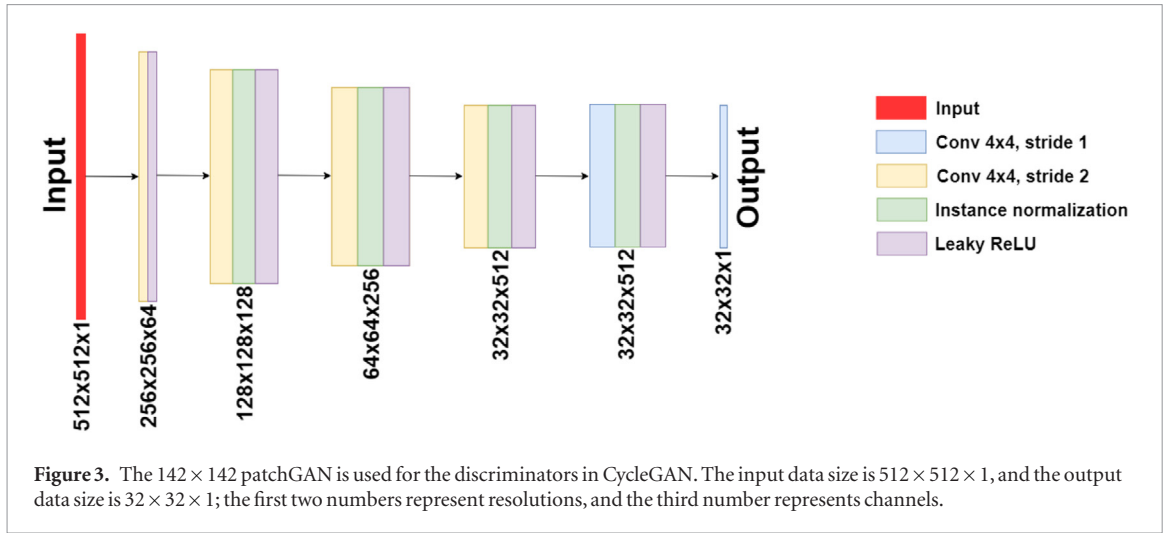
A network can still map images from one to multiple domains that share the same distribution features with adversarial loss only. Thus, to further reduce the mapping functions, cycle consistency loss is needed in this task, because it minimizes the differences between CT and cycle CT, and between CBCT and cycle CBCT. The cycle consistency losses for the two cycles are

$$\mathcal{L}_{Cycle-CT} = \frac{1}{m}\sum_{i=1}^{m} \left| G_A(G_B(CT_i)) - CT_i \right|, \tag{5}$$

and

$$\mathcal{L}_{Cycle-CBCT} = \frac{1}{m}\sum_{i=1}^{m} \left| G_B(G_A(CBCT_i)) - CBCT_i \right|. \tag{6}$$

To further preserve HU values between pCT and sCT, and between real CBCT and sCBCT, we add identity mapping loss in the loss function. G_A supposedly generates sCT from CBCT. However, if pCT images are input to G_A, the output is supposed to be pCT too and vice versa. Therefore, the identity mapping loss for CT and CBCT are expressed as follows:

Figure 3. The 142×142 patchGAN is used for the discriminators in CycleGAN. The input data size is $512 \times 512 \times 1$, and the output data size is $32 \times 32 \times 1$; the first two numbers represent resolutions, and the third number represents channels.

Figure 4. Illustration of a deformable H&N phantom. Area 1 represents the tumor, and area 2 represents the parotid gland.

$$\mathcal{L}_{Identity-CT} = \frac{1}{m}\sum_{i=1}^{m} |G_A(CT_i) - CT_i|, \tag{7}$$

$$\mathcal{L}_{Identity-CBCT} = \frac{1}{m}\sum_{i=1}^{m} |G_B(CBCT_i) - CBCT_i|. \tag{8}$$

Thus, combining all the losses in the previous equations, the full objective for the two generators is

$$\mathcal{L}_G = \mathcal{L}_{GAN-G_A} + \mathcal{L}_{GAN-G_B} + \alpha \times \left(\mathcal{L}_{Cycle-CT} + \mathcal{L}_{Cycle-CBCT}\right) + \beta \times \left(\mathcal{L}_{Identity-CT} + \mathcal{L}_{Identity-CBCT}\right), \tag{9}$$

where α is 10 and β is 5.

The full objective for the two discriminators is

$$\mathcal{L}_D = \mathcal{L}_{GAN-D_A} + \mathcal{L}_{GAN-D_B}. \tag{10}$$

2.5. Evaluation

In this experiment, we also tried another two unsupervised deep learning methods, DCGAN (Radford *et al* 2015) and PGGAN (Karras *et al* 2017), to train the models with the same training, validation, and test datasets.

For the patient study, dpCT images were used as the ground truth to evaluate the HU accuracy of sCT images generated by CycleGAN, DCGAN, and PGGAN. Evaluation was performed with similarity measures between sCT and dpCT images, including mean absolute error (MAE), root-mean-square error (RMSE), structural similarity index (SSIM), and peak-signal-to-noise ratio (PSNR). SSIM is calculated as $\mathrm{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$, where μ_x is the average of x, μ_y is the average of y, σ_x^2 is the variance of x, σ_y^2 is the variance of y, and σ_{xy} is the covariance of x and y. PSNR is calculated as $\mathrm{PSNR} = 20\log_{10}\frac{4071}{\sqrt{\mathrm{MSE}}}$. We also performed 3D gamma index analy-
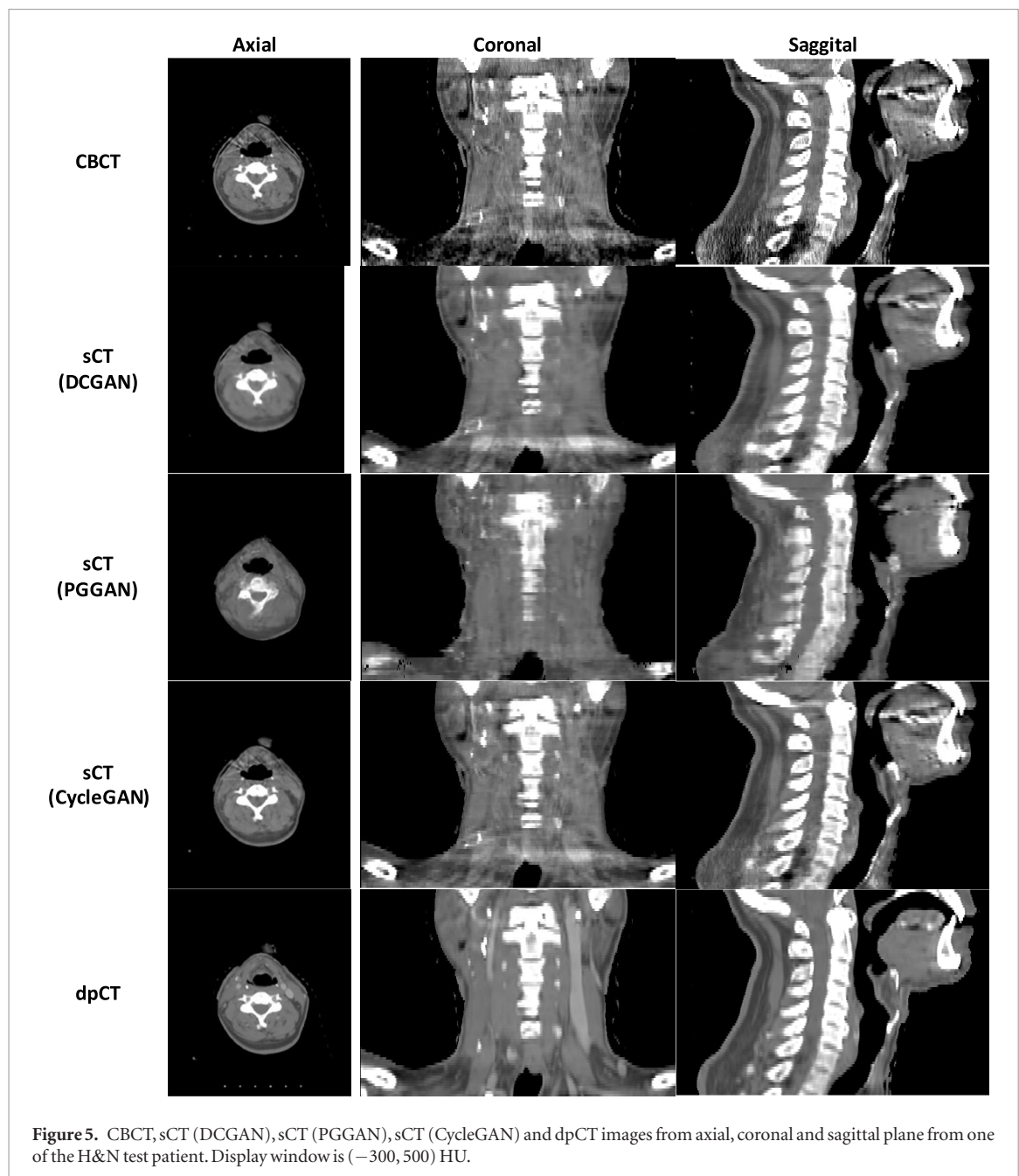
Figure 5. CBCT, sCT (DCGAN), sCT (PGGAN), sCT (CycleGAN) and dpCT images from axial, coronal and sagittal plane from one of the H&N test patient. Display window is $(-300, 500)$ HU.

sis, taking dpCT as reference, to show the HU accuracy of sCT images when used for dose calculation. Q–Q (quantile–quantile) plots between sCT and dpCT images and line profiles of HU value on images from samples of the test data were also plotted for detailed evaluation.

Real CT images, instead of dpCT images, are available as the ground truth in the phantom study. A deformable H&N phantom (Graves *et al* 2015) was used to compare the performance of DCGAN, PGGAN, CycleGAN and DIR. The phantom construction is shown in figure 4. In the phantom, the tumor (area1) and parotid gland (area2) can be deformed by injecting saline water and moving diode holders on a rail. We simulated tumor shrinkage and anatomy change by injecting 60 ml saline water into the phantom, then drawing 30 ml saline water out from the phantom and adjusting the diode holder positions. CT1 is acquired under 60 ml saline water, simulating pCT. CT2 and CBCT2 are acquired under 30 ml saline water, simulating the re-scanned CT and CBCT during the treatment course. sCT is generated by GAN models from CBCT2, and dpCT is generated by deforming CT1 to CBCT2's anatomy using the DIR package from Elastix. CT2 serves as the ground truth for quantitative evaluation.

3. Results

3.1. Scatter removal and HU improvement
Figure 5 shows CBCT, sCT generated by DCGAN, PGGAN and CycleGAN, and dpCT generated by DIR from one of the test patients. One can see that the CycleGAN model reduced scatter artifacts while retaining the anatomical
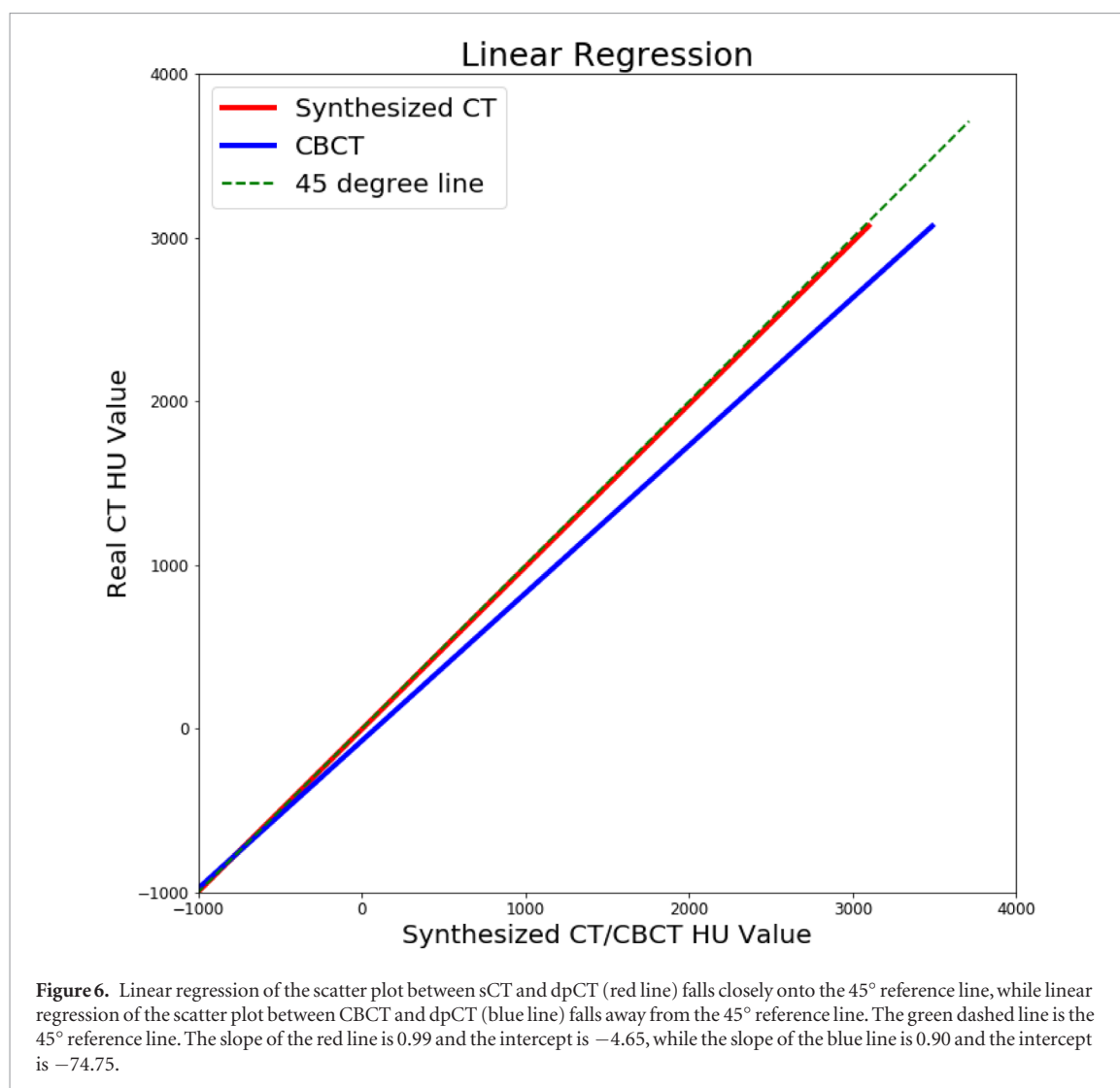
Figure 6. Linear regression of the scatter plot between sCT and dpCT (red line) falls closely onto the 45° reference line, while linear regression of the scatter plot between CBCT and dpCT (blue line) falls away from the 45° reference line. The green dashed line is the 45° reference line. The slope of the red line is 0.99 and the intercept is −4.65, while the slope of the blue line is 0.90 and the intercept is −74.75.

Table 1. Mean and standard deviation of similarity measures for CBCT volumes, sCT volumes (DCGAN), sCT volumes (PGGAN), and sCT volumes (CycleGAN) against dpCT volumes from all 20 test patients.

	CBCT	sCT (DCGAN)	sCT (PGGAN)	sCT (CycleGAN)
MAE (HU)	69.29 ± 11.01	39.71 ± 10.79	40.64 ± 6.99	**29.85 ± 4.94**
RMSE (HU)	182.8 ± 29.16	115.79 ± 27.47	120.45 ± 19.62	**84.46 ± 12.40**
SSIM	0.73 ± 0.04	0.81 ± 0.04	0.79 ± 0.03	**0.85 ± 0.03**
PSNR	25.28 ± 2.19	28.89 ± 2.16	28.65 ± 1.54	**30.65 ± 1.36**

accuracy when generating sCT images from the CBCT images. In contrast, the dpCT images have a slightly different anatomy distribution. The linear regression of the scatter plot between sCT (CycleGAN) and dpCT for all 20 test patients is shown in figure 6. If two datasets come from the same distribution, their linear regression plot should fall on the 45° reference line. The sCT and dpCT regression line is close to the 45° reference line, indicating that those two datasets are very similar, while the CBCT and dpCT regression line deviates from the 45° reference line, indicating that sCT greatly improves upon CBCT in terms of HU value accuracy. The results of similarity measures between CBCT volumes and dpCT volumes, and between sCT volumes and dpCT volumes for all 20 test patients are shown in table 1. sCT volumes from CycleGAN are more similar to dpCT volumes than sCT volumes from DCGAN and PGGAN.

Line profiles from one test patient are shown in figure 7. In the upper line profile, which passes through soft tissue and bone areas, sCT (CycleGAN) HU values are corrected to CT HU values. In the lower line profile, which passes only through the soft tissue area, CBCT HU values are noisy, while sCT (CycleGAN) HU values are smoothed and corrected to CT HU values. Q–Q plots between CBCT and dpCT, and between sCT (CycleGAN) and dpCT, from one test patient are shown in figure 8. The points of dpCT-sCT Q–Q plots falling along the 45° reference line regardless of soft tissue or bone areas indicates that sCT (CycleGAN) has the same distribution as dpCT.
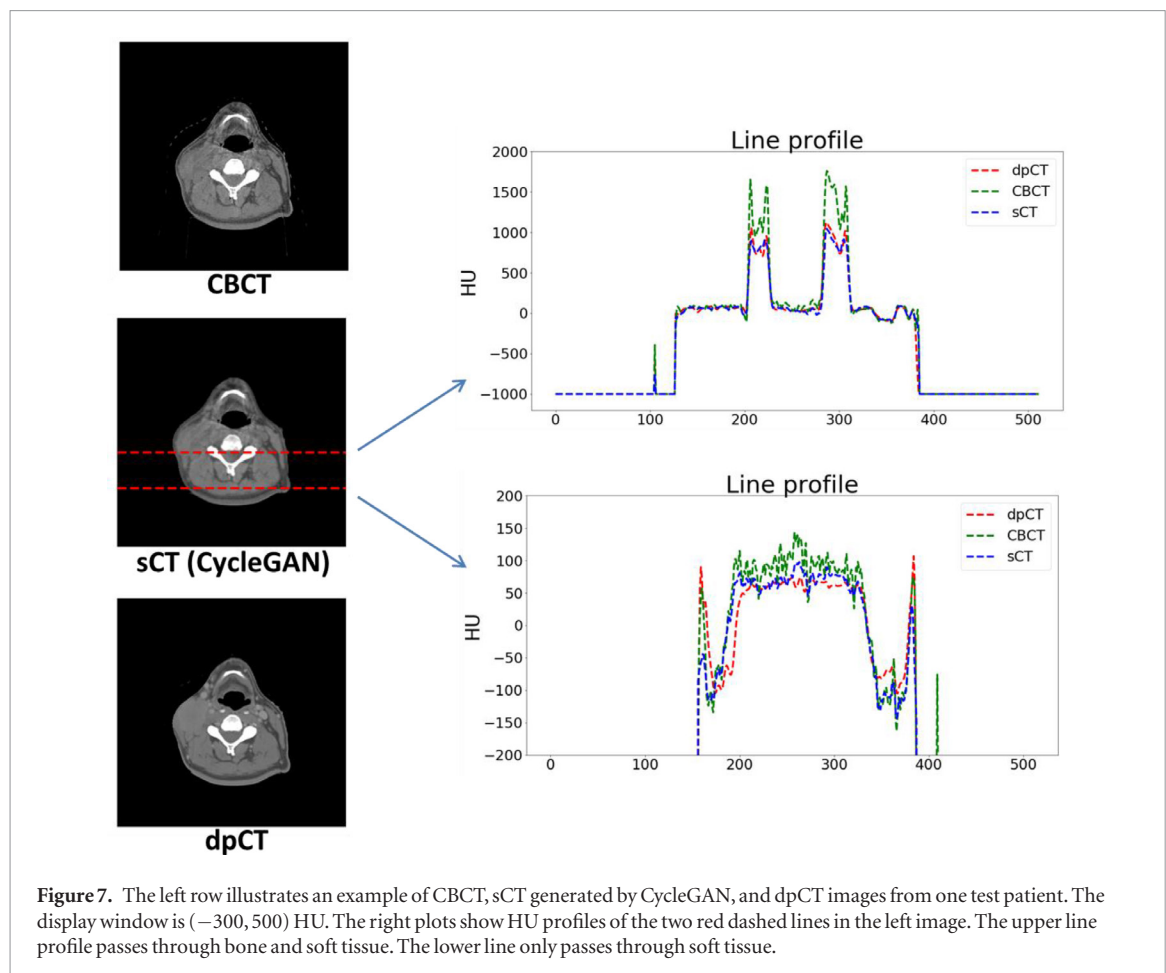
Figure 7. The left row illustrates an example of CBCT, sCT generated by CycleGAN, and dpCT images from one test patient. The display window is (−300, 500) HU. The right plots show HU profiles of the two red dashed lines in the left image. The upper line profile passes through bone and soft tissue. The lower line only passes through soft tissue.
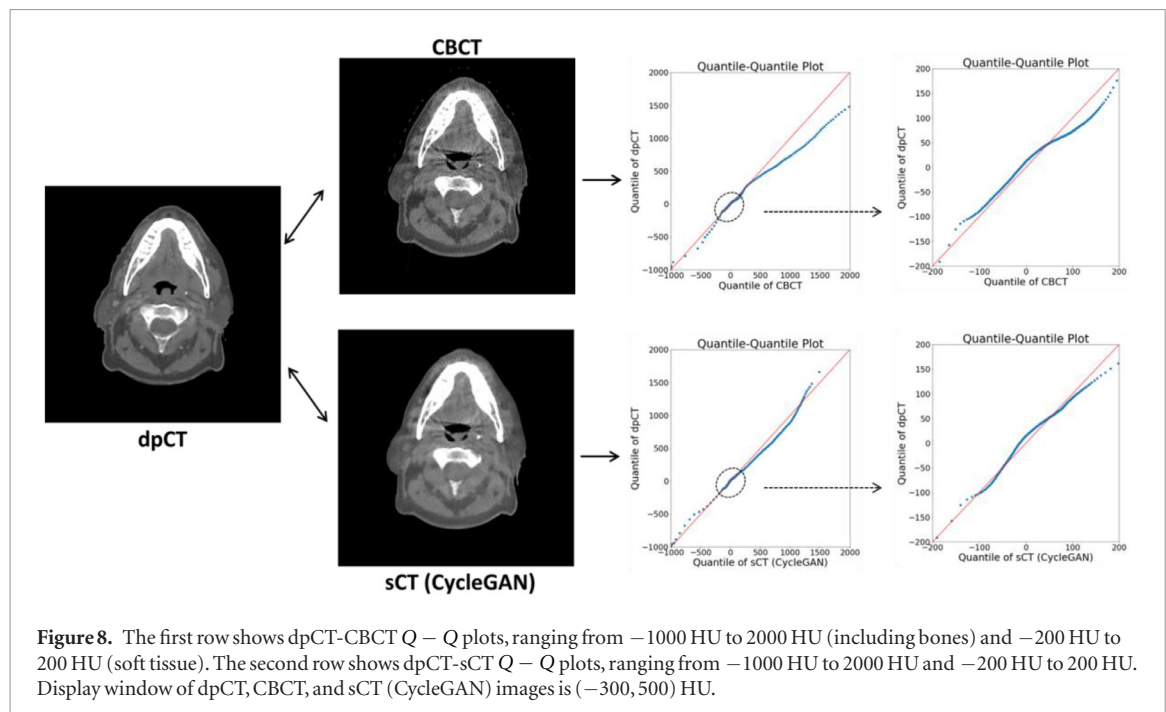

Figure 8. The first row shows dpCT-CBCT $Q-Q$ plots, ranging from −1000 HU to 2000 HU (including bones) and −200 HU to 200 HU (soft tissue). The second row shows dpCT-sCT $Q-Q$ plots, ranging from −1000 HU to 2000 HU and −200 HU to 200 HU. Display window of dpCT, CBCT, and sCT (CycleGAN) images is (−300, 500) HU.

3.2. Dose distribution comparison

Dose distributions on CBCT, sCT (DCGAN), sCT (PGGAN), sCT (CycleGAN) and dpCT were calculated with Eclipse treatment planning system by casting original clinical plans on them. 3D Gamma index pass rates for 2 mm/2% and 1 mm/1% criteria were calculated to evaluate sCT HU accuracy with dpCT as reference. All gamma index pass rates for CBCT and sCT from 20 test patients are plotted in figure 9, and their corresponding mean values and standard deviations are listed in table 2. Paired sample *t*-test was used to test CBCT and sCT (CycleGAN) gamma index pass rate for 2 mm/2% criteria. The test statistics are $t = 5.311$ and $p < 0.0001$ for a two-tailed test
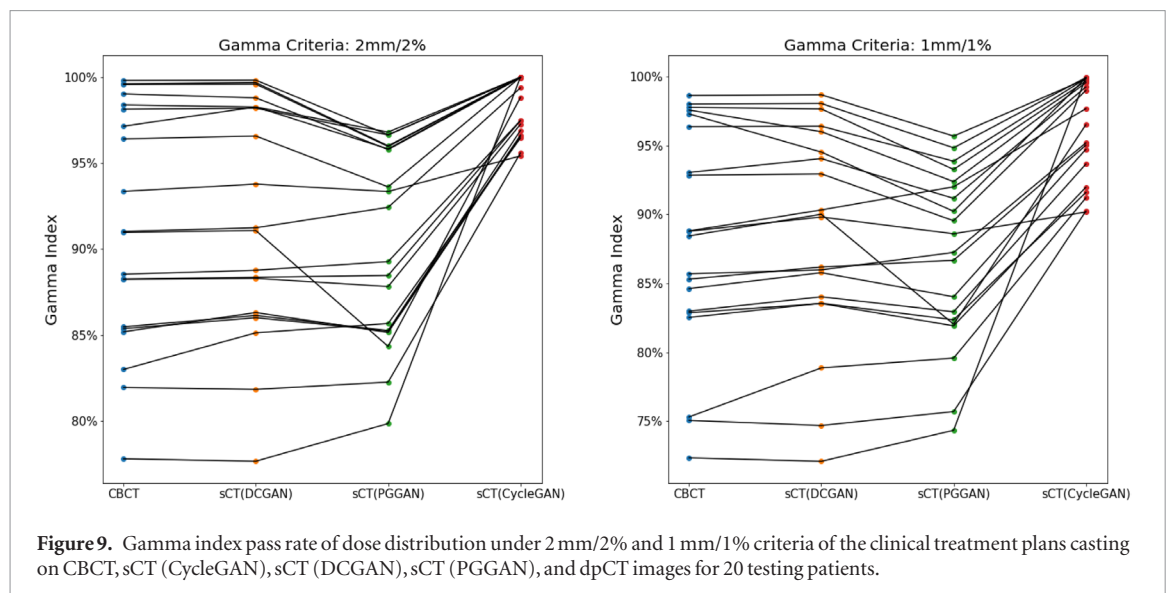
Figure 9. Gamma index pass rate of dose distribution under 2 mm/2% and 1 mm/1% criteria of the clinical treatment plans casting on CBCT, sCT (CycleGAN), sCT (DCGAN), sCT (PGGAN), and dpCT images for 20 testing patients.

Table 2. 3D gamma analysis of dose distributions on CBCT and sCTs generated by different GAN models compared with dose distribution on dpCT for the clinical treatment plans from 20 testing patients. The numbers are mean and standard deviation of gamma index pass rate under 2 mm/2% and 1 mm/1% criteria.

	CBCT	sCT (DCGAN)	sCT (PGGAN)	sCT (CycleGAN)
Gamma criteria: 2 mm/2%	91.37% ± 6.72%	91.69% ± 6.56%	90.32% ± 5.44%	**98.40% ± 1.68%**
Gamma criteria: 1 mm/1%	88.22% ± 8.02%	88.66% ± 7.53%	86.92% ± 6.14%	**96.26% ± 3.59%**

with $\alpha = 0.05$ and $df = 19$. Therefore, there is enough evidence to claim that the population mean of the CBCT gamma index pass rate is different than the sCT (CycleGAN) gamma index pass rate at the 0.05 significance level. Figure 10 shows dose calculation results of the clinical treatment plan from one of the test patients. Visually, the dose distributions of sCT (CycleGAN) are closer to those of dpCT than the others. Quantitatively, by 3D gamma analysis, dose distributions on sCT (CycleGAN) have higher gamma pass rates with mean values of 96.26% and 98.40%, and standard deviations of 3.59% and 1.68% under 1 mm/1% and 2 mm/2% criteria, respectively.

3.3. Phantom study

The images of CT1, CBCT2, sCT (CycleGAN), sCT (DCGAN), sCT (PGGAN), dpCT, and CT2 from the deformable phantom are shown in figure 11. sCT (CycleGAN) and sCT (DCAGN) kept the same anatomy as CT2, while at the same time, scatter artifacts were removed. sCT (PGGAN) had the worst performance and generated fake images. When comparing dpCT against CT2 in this phantom study, the DIR method performed well in terms of capturing the soft tissue change. Table 3 shows the results of similarity measures of CT1, CBCT2, sCT (DCGAN), sCT (PGGAN), sCT (CycleGAN) and dpCT with ground truth CT2. CycleGAN model had the best performance among the GANs as well as the DIR method in this case, both visually and quantitatively.

4. Discussion and conclusions

We developed a CycleGAN model to convert CBCT to CT-like images. In the patient study, dpCT images were used as ground truth to evaluate HU accuracy of sCT. Even though dpCT has minor anatomical accuracy issues, dpCT is the best approximation to the ground truth in this problem space, as it keeps the same HU accuracy as the real CT. Therefore, it is reasonable to use dpCT to evaluate sCT's HU accuracy. Visually and quantitatively, sCT (CycleGAN) is more similar to dpCT than DCGAN and PGGAN in terms of HU similarity measures and gamma index pass rate. The loss functions in the DCGAN and PGGAN model only include the adversarial loss, and the cycle consistency loss used in CycleGAN is not applicable to these two particular frameworks. A model trained using only an adversarial loss has been shown to be able to generate synthetic data that has a distribution that matches that of the target data (Goodfellow *et al* 2014), creating images that look similar to the target images. However, the model may fall short in producing images with particular aspects essential to medical data, such as high anatomical accuracy. In the DCGAN (Radford *et al* 2015) and PGGAN (Karras *et al* 2017) papers, they both pointed about the model instability. The additional cycle consistency loss from CycleGAN addresses this issue. In the phantom study, the real CT images were used as the ground truth to compare DCGAN, PGGAN, CycleGAN and the DIR method. CycleGAN outperformed DCGAN, PGGAN, and the DIR method in similarity measures, while DCGAN kept more scatter artifacts from CBCT, PGGAN generated distorted images, and DIR had inaccurate soft tissue deformation.
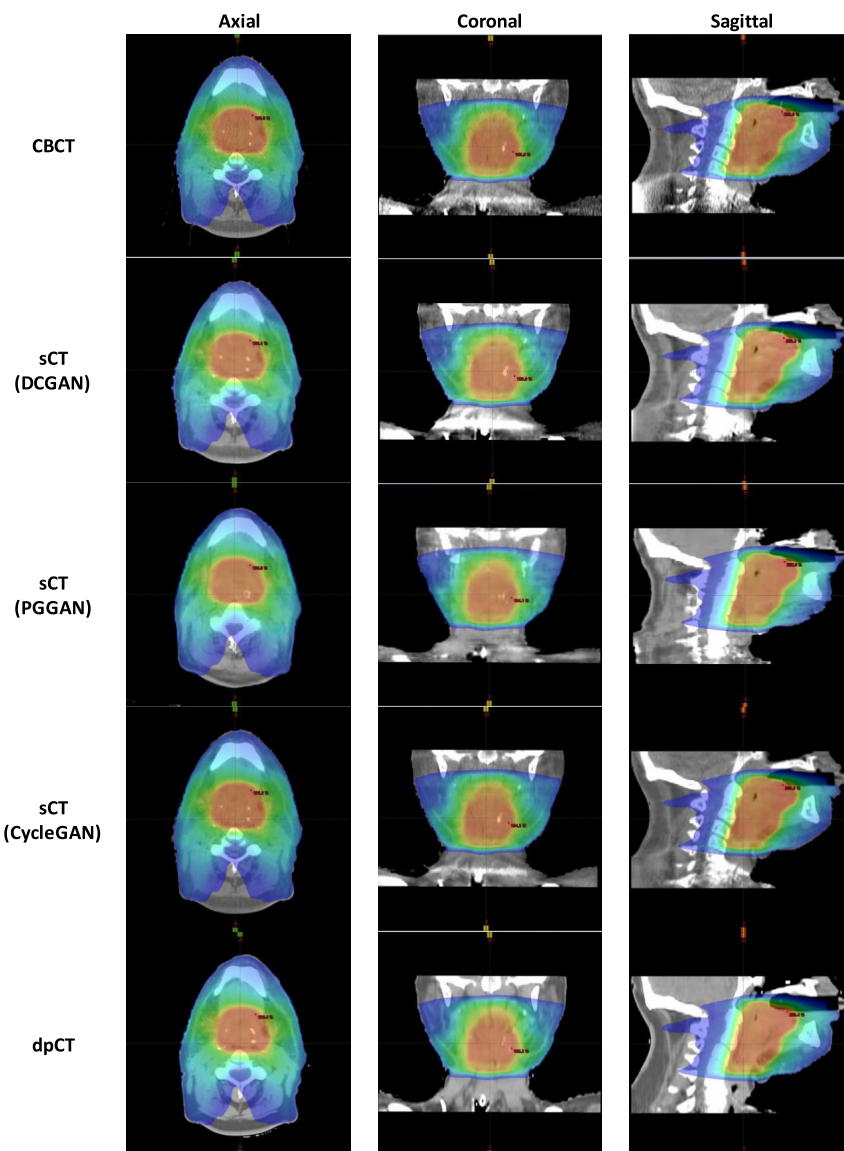
Figure 10. Dose distributions of the clinical treatment plan casting on CBCT, sCT (DCGAN), sCT (PGGAN), sCT (CycleGAN), and dpCT images for one of the testing patients.

Table 3. Similarity measures for CT1, CBCT2, sCT (DCGAN), sCT (PGGAN), sCT (CycleGAN), and dpCT against CT2 in the phantom study.

	CT1	CBCT2	sCT (DCGAN)	sCT (PGGAN)	sCT (CycleGAN)	dpCT
MAE (HU)	8.43	28.54	11.38	30.17	**5.02**	6.98
RMSE (HU)	28.40	59.46	26.26	70.48	**16.87**	22.97
SSIM	0.89	0.73	0.84	0.72	**0.93**	0.91
PSNR	27.22	21.06	27.87	21.39	**31.85**	29.75

For the CycleGAN model, unpaired CBCT and CT images can be used for training to avoid the requirement of paired CBCT and CT datasets that are difficult to acquire in reality. Once training is completed, generating sCT images from CBCT by this model can be accomplished in seconds. Furthermore, through the CycleGAN model, most scatter artifacts in CBCT are removed, and the HU values are corrected. Thus, CycleGAN can be effectively used for dose calculation in ART.

One concern about CycleGAN is its application range. The training and testing images in this experiment are only randomly partitioned from a uniform acquisition cohort. It is possible to use the same model on different cohorts from other institutions by adding more data from other cohorts into the training dataset. It is also possible to use the same model on entirely different anatomies by adding more images from other sites into the training dataset. To demonstrate this point, figure A1 in the appendix shows sCT generated from the same CycleGAN model for a pelvic patient and the performance of the CycleGAN model on pelvic site is shown in table A1.
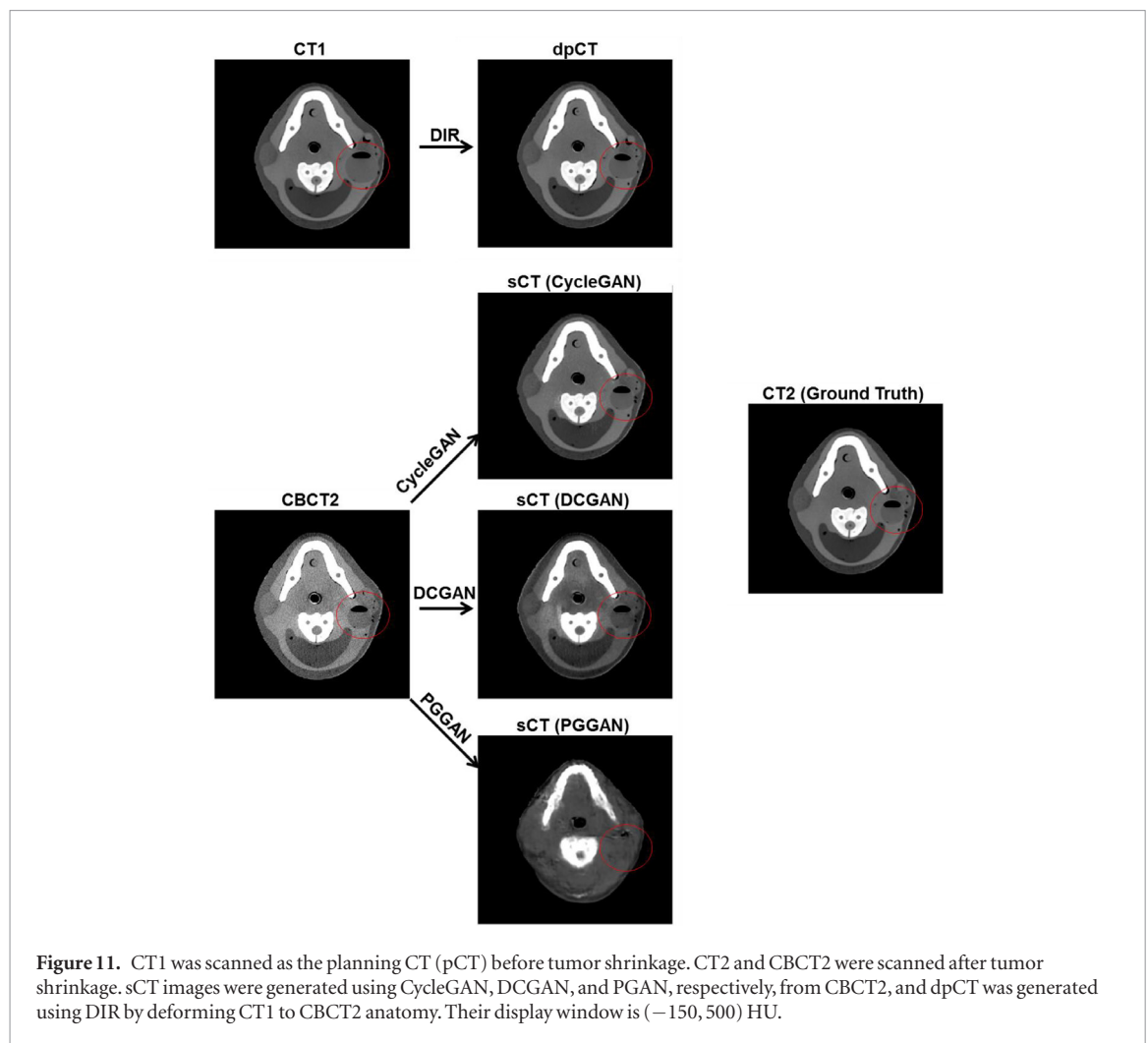
Figure 11. CT1 was scanned as the planning CT (pCT) before tumor shrinkage. CT2 and CBCT2 were scanned after tumor shrinkage. sCT images were generated using CycleGAN, DCGAN, and PGAN, respectively, from CBCT2, and dpCT was generated using DIR by deforming CT1 to CBCT2 anatomy. Their display window is (−150, 500) HU.

Another concern is the training stability of CycleGAN. It is known that the stability of different GAN models is highly dependent on network architecture, loss functions, training schedules, etc. Therefore, an unbalanced generator and discriminator design and training schedule will lead to instability and catastrophic failure modes. Image size is one of the most important factors that will affect the generator and discriminator design. High resolution images generally require deeper networks, since shallow networks cannot capture global feature on a large image size. A CycleGAN model that is designed and trained to work well on a 256×256 resolution image may not have good performance on a 512×512 resolution image. A balance needs to be found between generator and discriminator to get a stable CycleGAN model. In our experiment, the loss of both training and validation of the CycleGAN were designed for the convergence of images at 512×512 resolution.

Even though CycleGAN can remove most scatter artifacts on the CBCT images and correct their HU values to match those of the CT images, it is unable to reduce metal artifacts, since both the CT and CBCT training datasets include metal artifacts. Motion artifacts are not severe in our dataset due to the static anatomical sites we studied in this work. We intend to investigate the management of such artifacts in future studies.

In our training and test datasets, some CBCT images have severe truncation problems, yielding generated sCT images that also have the truncations. PGGAN tries to add new anatomy on the truncated anatomy. However, it generates false anatomy, causing poor performance in our experiment. It is also very time consuming and computationally intensive during training, due to its multi-resolution learning approach.

As a future study, we plan to investigate solutions to the image truncation problem caused by the small field of view. Successful completion of such work would allow for accurate and efficient ART planning on sCT images generated from CBCT images. In addition, we plan to add data from other anatomical sites into the development of the CycleGAN model, improving its stability and generalizability.

Acknowledgment

Appendix

Figure A1. CBCT, sCT generated from CycleGAN and dpCT images from axial, coronal and sagittal plane around pelvic region. Display window is $(-400, 400)$ HU.
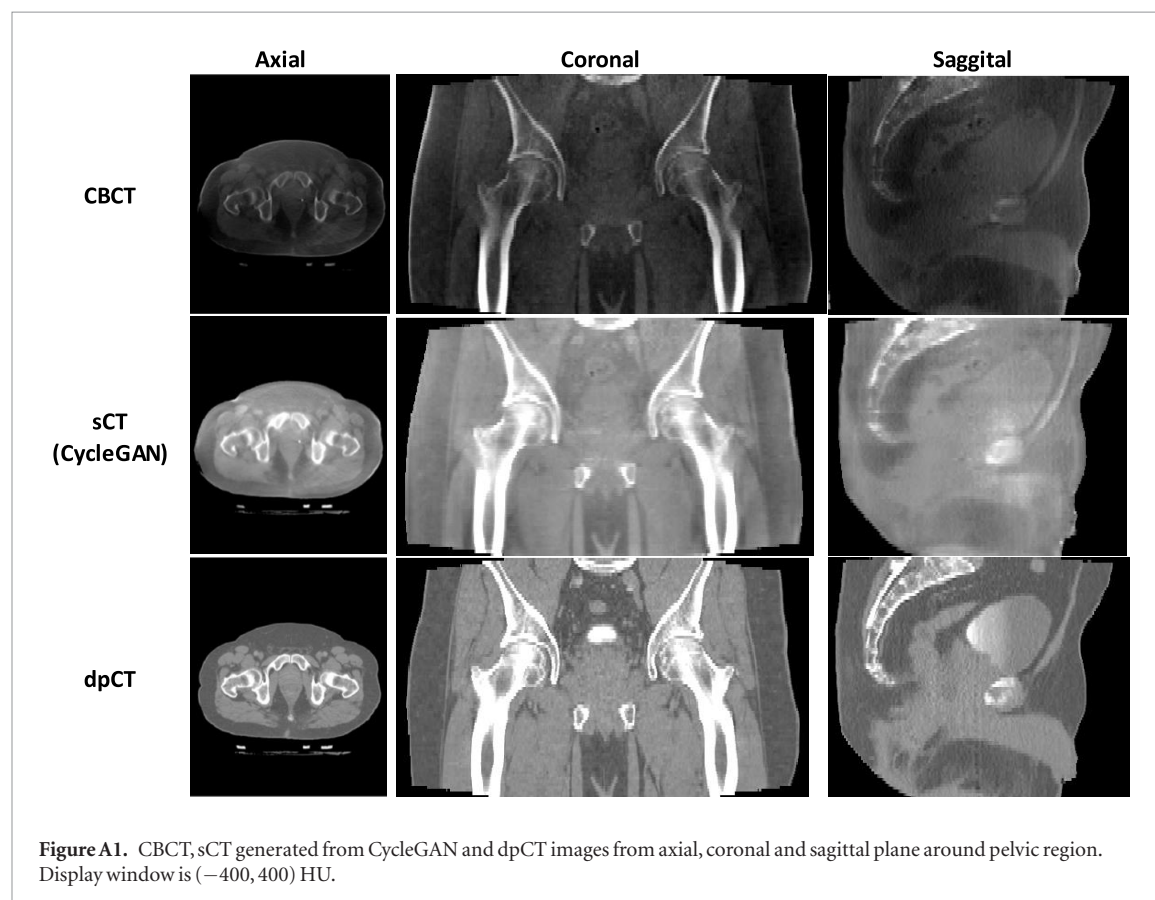
Table A1. Mean and standard deviation of similarity measures for CBCT volumes and sCT volumes (CycleGAN) against dpCT volumes from 20 pelvic cancer patients.

	CBCT	sCT (CycleGAN)
MAE (HU)	92.83 ± 14.66	$\mathbf{38.90 \pm 9.47}$
RMSE (HU)	178.03 ± 20.64	$\mathbf{98.23 \pm 18.76}$
SSIM	0.63 ± 0.08	$\mathbf{0.75 \pm 0.06}$
PSNR	23.39 ± 1.51	$\mathbf{28.97 \pm 1.21}$

ORCID iDs

Xiao Liang ● https://orcid.org/0000-0002-2472-2396
Liyuan Chen ● https://orcid.org/0000-0002-5345-0086
Dan Nguyen ● https://orcid.org/0000-0002-9590-0655
Xuejun Gu ● https://orcid.org/0000-0001-6725-6906
Ming Yang ● https://orcid.org/0000-0001-9122-0401

References

Abe T, Tateoka K, Saito Y, Nakazawa T, Yano M, Nakata K, Someya M, Hori M and Sakata K 2017 Method for converting cone-beam CT values into Hounsfield units for radiation treatment planning *Int. J. Med. Phys. Clin. Eng. Radiat. Oncol.* **6** 361
Barker J L *et al* 2004 Quantification of volumetric and geometric changes occurring during fractionated radiotherapy for head-and-neck cancer using an integrated CT/linear accelerator system *Int. J. Radiat. Oncol. Biol. Phys.* **59** 960–70
Gan Z, Chen L, Wang W, Pu Y, Zhang Y, Liu H, Li C and Carin L 2017 Triangle generative adversarial networks *Proc. of the 31st Int. Conf. on Neural Information Processing Systems (Long Beach, California, USA: Curran Associates Inc.)* pp 5253–62
Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A and Bengio Y 2014 Generative adversarial nets *Proc. of the 27th Int. Conf. on Neural Information Processing Systems* vol 2 (Montreal, Canada: MIT Press) pp 2672–80
Graves Y J, Smith A A, Mcilvena D, Manilay Z, Lai Y K, Rice R, Mell L, Jia X, Jiang S B and Cerviño L 2015 A deformable head and neck phantom with *in vivo* dosimetry for adaptive radiotherapy quality assurance *Med. Phys.* **42** 1490–7

Hansen E K, Bucci M K, Quivey J M, Weinberg V and Xia P 2006 Repeat CT imaging and replanning during the course of IMRT for head-and-neck cancer *Int. J. Radiat. Oncol. Biol. Phys.* **64** 355–62

Hiasa Y, Otake Y, Takao M, Matsuoka T, Takashima K, Carass A, Prince J L, Sugano N and Sato Y 2018 Cross-Modality Image Synthesis from Unpaired Data Using CycleGAN *(Simulation and Synthesis in Medical Imaging, SASHIMI 2018) Lecture Notes in Computer Science* vol 11037 ed Gooya A, Goksel O, Oguz I, Burgos N (Berlin: Springer) pp 31–41

Isola P, Zhu J-Y, Zhou T and Efros A A 2017 Image-to-image translation with conditional adversarial networks *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* pp 5967–76

Karras T, Aila T, Laine S and Lehtinen J 2017 Progressive growing of GANs for improved quality, stability, and variation (arXiv:1710.10196)

Klein S, Staring M, Murphy K, Viergever M A and Pluim J P 2010 Elastix: a toolbox for intensity-based medical image registration *IEEE Trans. Med. Imaging* **29** 196–205

Kurz C, Kamp F, Park Y K, Zöllner C, Rit S, Hansen D, Podesta M, Sharp G C, Li M and Reiner M 2016 Investigating deformable image registration and scatter correction for CBCT-based dose calculation in adaptive IMPT *Med. Phys.* **43** 5635–46

Lei Y, Tang X, Higgins K, Wang T, Liu T, Dhabaan A, Shim H, Curran W J and Yang X 2018 Improving image quality of cone-beam CT using alternating regression forest *Proc. SPIE* **10573** 1057345

Miracle A and Mukherji S 2009 Conebeam CT of the head and neck, part 1: physical principles *Am. J. Neuroradiol.* **30** 1088–95

Naimuddin S, Hasegawa B and Mistretta C A 1987 Scatter-glare correction using a convolution algorithm with variable weighting *Med. Phys.* **14** 330–4

Radford A, Metz L and Chintala S 2015 Unsupervised representation learning with deep convolutional generative adversarial networks (arXiv:1511.06434)

Schwartz D L, Garden A S, Shah S J, Chronowski G, Sejpal S, Rosenthal D I, Chen Y, Zhang Y, Zhang L and Wong P-F 2013 Adaptive radiotherapy for head and neck cancer—dosimetric results from a prospective clinical trial *Radiother. Oncol.* **106** 80–4

Wang T-C, Liu M-Y, Zhu J-Y, Tao A, Kautz J and Catanzaro B 2018 High-resolution image synthesis and synthetic manipulation with conditional GANs *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* (IEEE) pp 8798–807

Wolterink J M, Dinkla A M, Savenije M H, Seevinck P R, van den Berg C A and Išgum I 2017 Deep MR to CT synthesis using unpaired data *Simulation and Synthesis in Medical Imaging* (Berlin: Springer) pp 14–23

Xu Y, Bai T, Yan H, Ouyang L, Pompos A, Wang J, Zhou L, Jiang S B and Jia X 2015 A practical cone-beam CT scatter correction method with optimized Monte Carlo simulations for image-guided radiation therapy *Phys. Med. Biol.* **60** 3567

You S H, Kim S Y, Lee C G, Keum K C, Kim J H, Lee I J, Kim Y B, Koom W S, Cho J and Kim S K 2012 Is there a clinical benefit to adaptive planning during tomotherapy in patients with head and neck cancer at risk for xerostomia? *Am. J. Clin. Oncol.* **35** 261–6

Zhu J-Y, Park T, Isola P and Efros A A 2017 Unpaired image-to-image translation using cycle-consistent adversarial networks *2017 IEEE International Conference on Computer Vision (ICCV)* pp 2242–51

Zöllner C, Rit S, Kurz C, Vilches-Freixas G, Kamp F, Dedes G, Belka C, Parodi K and Landry G 2017 Decomposing a prior-CT-based cone-beam CT projection correction algorithm into scatter and beam hardening components *Phys. Imaging Radiat. Oncol.* **3** 49–52