



Published in final edited form as:

*Phys Med Biol.* ; 64(14): 145015. doi:10.1088/1361-6560/ab25bc.

## MRI-based Treatment Planning for Proton Radiotherapy: Dosimetric Validation of a Deep Learning-based Liver Synthetic CT Generation Method

Yingzi Liu<sup>1</sup>, Yang Lei<sup>1</sup>, Yinan Wang<sup>1</sup>, Tonghe Wang<sup>1</sup>, Lei Ren<sup>2</sup>, Liyong Lin<sup>1</sup>, Mark McDonald<sup>1</sup>, Walter J. Curran<sup>1</sup>, Tian Liu<sup>1</sup>, Jun Zhou<sup>1</sup>, and Xiaofeng Yang<sup>1</sup>

<sup>1</sup>Department of Radiation Oncology and Winship Cancer Institute, Emory University, Atlanta, GA 30322

<sup>2</sup>Department of Radiation Oncology, Duke University, Durham, NC 27708

### Abstract

Magnetic resonance imaging (MRI) has been widely used in combination with computed tomography (CT) radiation therapy because MRI improves the accuracy and reliability of target delineation due to its superior soft tissue contrast over CT. The MRI-only treatment process is currently an active field of research since it could eliminate systematic MR-CT co-registration errors, reduce medical cost, avoid diagnostic radiation exposure, and simplify clinical workflow. The purpose of this work is to validate the application of a deep learning-based method for abdominal synthetic CT (sCT) generation by image evaluation and dosimetric assessment in a commercial proton pencil beam treatment planning system (TPS). This study proposes to integrate dense block into a 3D cycle-consistent generative adversarial networks (cycle GAN) framework in an effort to effectively learn the nonlinear mapping between MRI and CT pairs. A cohort of 21 patients with co-registered CT and MR pairs were used to test the deep learning-based sCT image quality by leave-one-out cross validation. The CT image quality, dosimetric accuracy and the distal range fidelity were rigorously checked, using side-by-side comparison against the corresponding original CT images. The average mean absolute error (MAE) was  $72.87 \pm 18.16$  HU. The relative differences of the statistics of the PTV dose volume histogram (DVH) metrics between sCT and CT were generally less than 1%. Mean 3D gamma analysis passing rate of 1mm/1%, 2mm/2%, 3mm/3% criteria with 10% dose threshold were  $90.76 \pm 5.94\%$ ,  $96.98 \pm 2.93\%$  and  $99.37 \pm 0.99\%$ , respectively. The median, mean and standard deviation of absolute maximum range differences were 0.170 cm, 0.186 cm and 0.155 cm. The image similarity, dosimetric and distal range agreement between sCT and original CT suggests the feasibility of further development of an MRI-only workflow for liver proton radiotherapy.

---

After the embargo period, everyone is permitted to use copy and redistribute this article for non-commercial purposes only, provided that they adhere to all the terms of the licence <https://creativecommons.org/licenses/by-nc-nd/3.0>

**Corresponding author:** Xiaofeng Yang, PhD, Department of Radiation Oncology, Emory University School of Medicine, 1365 Clifton Road NE, Atlanta, GA 30322, Tel: (404)-778-8622, Fax: (404)-778-4139, [xyang43@emory.edu](mailto:xyang43@emory.edu).

**Publisher's Disclaimer:** Accepted Manuscript is "the version of the article accepted for publication including all changes made as a result of the peer review process, and which may also include the addition to the article by IOP Publishing of a header, an article ID, a cover sheet and/or an 'Accepted Manuscript' watermark, but excluding any other editing, typesetting or other changes made by IOP Publishing and/or its licensors"

## 1. INTRODUCTION

Worldwide, approximately 50% of all cancer patients undergo radiotherapy (Delaney *et al.*, 2005). As a highly advanced form of radiotherapy, proton therapy offers an important advantage over photons in terms of the depth dose distribution. It provides a sharp dose fall-off beyond the target which spares the normal tissues from radiation (Levin *et al.*, 2005). The recent adoption of pencil beam scanning (PBS) technique allows proton beam to be delivered spot-by-spot with modulation of intensity, lateral scanning position and penetration depth (Kanai *et al.*, 1980). PBS also dramatically decreases the neutron dose that produced by the modulator and collimator from the double scattering technique. Like photon radiotherapy, the current proton treatment planning is depended on computed tomography (CT). CT is currently the clinically used image modality that provides the electron density information that is necessary for dose calculation and digitally reconstructed radiograph (DRR) generation. To precisely and robustly delineate target structures and organs at risk (OARs), magnetic resonance imaging (MRI) is often used as a complementary modality to CT in radiotherapy. Transformation of the contours from MR to CT necessitates the registration between MR and CT images, which can introduce an undesirable 2–5 mm systematic error (Edmund and Nyholm, 2017; Roberson *et al.*, 2005; Ulin *et al.*, 2010; Dean *et al.*, 2012; Daisne *et al.*, 2003; Nyholm *et al.*, 2009), leading to a geometric miss and compromised PTV margin (Edmund and Nyholm, 2017). The motion induced movement of tumors has posed a major concern for treating lung and liver cancer and other locations in the thorax and abdomen. The differential movement of the primary tumor and lymph nodes is not only occurring at inter- but also intra-fractional radiotherapy (De Ruyscher *et al.*, 2015). The clinical introduction of MR-guided radiotherapy has allowed the mitigation or correction of motion artifact (Lagendijk *et al.*, 2014; Kontaxis *et al.*, 2017; Oborn *et al.*, 2017a), and is paving the way toward on-line adaptive radiotherapy. Motivated by eliminating systematic error and the emerging MRI imaging guidance in radiotherapy, MR-only treatment planning has become an active field of research, in which MRI can be used as the sole imaging modality. MR-only treatment workflow can also spare the patient from CT radiation doses, which benefit more for pediatric patient that have much less dose upper limit (Dougeni *et al.*, 2012) and for those patients that image-guided radiotherapy where multiple cone-beam CTs (CBCTs) are acquired (Wen *et al.*, 2007). One major task in any MR-only treatment workflow is the generation of synthetic CT (sCT) images. These images can then serve as CT surrogates that can be used for dose calculation and digital reconstructed radiograph generation. Since the physics of X-ray and proton interactions in matter are fundamentally different, proton dose calculated is more sensitive to the local mismatch and HU accuracy. Therefore, the proton dosimetric results are supposed to be quite different among different sCT generation methods, which enables a direct comparison of the superiority.

The currently available methods to produce sCT broadly fall into the following three categories: segmentation-based (Chin *et al.*, 2014; Korhonen *et al.*, 2014; Bredfeldt *et al.*, 2017; Hsu *et al.*, 2013), atlas-based (Sjölund *et al.*, 2015; Guerreiro *et al.*, 2017; Lei *et al.*, 2019a; Lei *et al.*, 2018a) and machine learning-based methods (Han, 2017; Lei *et al.*, 2018b; Yang *et al.*, 2019). The tissue HU prediction of both segmentation and atlas-based methods

depends on either the predetermined values or the atlas CT number rather than patient-specific HU values derived from learning-based methods. Learning-based methods can be further broken into different categories such as the random forests and deep-learning. The key difference between deep-learning and random forest is that the former can automatically learn useful features of the data, eliminating the need for handcrafted features such as Haar-like and discrete cosine transform (DCT) used in random forest methods (Huynh *et al.*, 2016). In deep-learning-based methods, convolutional neural networks (CNNs) were introduced by Li *et al.* to generate a PET attenuation correction map. One limitation of CNN-based method is that it can produce blurry results due to MR-CT local mismatch. Recently, generative adversarial networks (GANs) have been proposed by incorporating an adversarial loss term to produce more realistic sCT (Nie *et al.*, 2017; Emami *et al.*, 2018). However, GAN-based methods still require the MR-CT pairs to be perfectly registered, which can be difficult especially in the body sites like abdomen.

Recently, we proposed a novel deep learning-based algorithm based on a 3D cycle GAN to generate MRI-based sCT (Lei *et al.*, 2019c). This work aimed to apply this method to generate abdominal sCT for MRI-based proton radiotherapy. At the site of abdomen, the image quality is commonly affected by intrinsic organ motion, which can lead to significant artifacts without motion control. These artifacts make the accurate sCT prediction particularly difficult. In addition, to treat the target in liver, the proton beams usually have to go through the small rib bones, which are rather challenging to generate in sCT. In our deep learning-based method, a novel 3D cycle-consistent GAN with integrated dense block minimization to capture 3D spatial information and to cope with local mismatches between MR and CT paired images. To better differentiate bone from air structure and to retain sCT image sharpness, a novel compound loss function was employed in the architecture. To explore whether the sCT can be robust used for proton treatment planning, evaluation of the dosimetric and the distal range agreement between the sCT and the original CT was carried out.

## 2. MATERIALS AND METHODS

### 2.A. Image acquisition

The study cohort was composed of 21 patients diagnosed with hepatocellular carcinoma and originally treated with liver photon SBRT. The cancer stage was either T1N0M0 or T2N0M0. The patient age varied from 50 to 80. Treatment time ranged from 2010 to 2018. Image data were extracted retrospectively under an IRB-approved protocol. Routine abdominal CT and MR scans were acquired on the same day with either breath-hold (16 patients) or abdominal compression (5 patients) to minimize the respiratory motion. CT scans were acquired on a Siemens (Erlangen, Germany) Biograph40. The CT acquisition parameters were: 120 kVp, 1.523 mm × 1.523 mm × 2 mm voxel size, and 780 mm × 780 mm field-of-view (FOV), and the acquisition length in axial direction ranges from 160 mm to 598 mm. T1-weighted MRIs were acquired on Siemens Biograph40 3T, Siemens TrioTim 3T, and GE Signa HDxt 1.5T. 3D fat-suppressed fast field echo images were acquired at the Siemens Skyra (2 patients) and Siemens TrioTim (2 patients) using volumetric interpolated breath-hold examination (VIBE). The sequence applied for these two methods were TE/TR

= 1.34/4.34 ms and TE/TR = 2.45, 2.46/5.27, 6.47 ms, respectively. FOVs were 440 mm × 275 mm and 440 mm × 288.75 mm, respectively, and the acquisition length in axial direction was 288 mm. The flip angles were both 9 degree. The voxel sizes were both 1.375 mm × 1.375 mm × 3 mm. 2D fat-suppressed fast spoiled gradient echo was applied at the GE Signa HDxt for the rest of the 17 patients. The sequence parameters were: TE ranging from 2.2 to 4.4 ms, TR ranging from 175 to 200 ms, patient position being FFS and flip angle being 80 degree. FOV was 480 mm × 480 mm and the acquisition length in axial direction was from 117 mm to 300 mm. The voxel size was 1.875 mm × 1.875 mm × 3 mm. Geometrical correction was performed using the built-in software package at the scanner. Respiratory belt was used to monitor breathing for the breath-hold scans. The scanning time for the T1 imaging was 20–26 seconds.

Anatomical structures were contoured by physicians for treatment planning.

## 2.B. Image pre-processing and registration

First, the intensity inhomogeneity of the MR images was corrected by the N4ITK MRI Bias correction filter, available at the open source 3D SLICER 4.8.1. N4ITK was used with BSpline grid resolution of 10,10,10, and the other parameters were the default values. The MR images were then rigidly registered and deformed to match with the corresponding CT images using Velocity AI 3.2.1 (Varian Medical Systems, Inc. Palo Alto, USA). The option of MR corrected deformable was used as the algorithm to deform the MR images to the CT images. Resample was applied on the deformed MR images. Finally, the registered MR images and their CT pairs were uploaded our machine-learning algorithm to train.

## 2.C. sCT generation

For the cohort of 21 patients, we used leave-one-out cross-validation. Given the degree of organ motion and the complexity of the required registrations, application of traditional convolution neural networks (CNN) to generate abdominal sCT could lead to errors (Wolterink *et al.*, 2017). To overcome this, we used a novel 3D cycle GAN that contains several dense blocks in the generator to capture both the structural and textural information and to cope with local mismatches between MR and CT images. Compared to CT, MR images have more structural information and contrast in soft tissue regions and less at bone and air interfaces. The traditional MR-to-CT GAN is thus bound to generate erroneous prediction where the many-to-one or one-to-many mapping happens. To deal with this issue, we applied an inverse MRI-to-CT transformation model by incorporating “cycle GAN” (Zhu *et al.*, 2017) to approach one-to-one MR to CT mapping. To solve the problem of possible cross-slice discontinuousness (Largent *et al.*, 2018), a 3D image patch (voxel size [64, 64, 5]) was adopted as the input of this model. MRI and CT are essentially two different image modalities, dense blocks were therefore employed to combine low and high frequency information to effectively represent image patches between these two. As is shown in the generator architecture in Figure 1, the feature map first undergoes two down-sampling convolutional layers to be downsized, then it passes 9 dense blocks, after which it goes through two deconvolutional layers and a tanh layer to enable an end-to-end mapping. The tanh layer works as a nonlinear activation function which facilitates the model to generalize or adapt to a variety of data that can differentiate the outputs, for example determining

whether a voxel on a boundary is bone or air. As shown in Figure 1, each dense block is implemented by six convolution layers. The first convolution layer is applied to the input to create  $k$  feature maps. The following four layers are applied to the concatenated information of all the previous feature maps and input to create more feature maps in sequence. The final output of these layers thus contains  $5*k$  feature maps. Finally, the output goes through the last layer to shorten the feature maps to  $k$ . The low frequency signal that contains the texture information is obtained from former convolutional layers. The high frequency signal that contains the structural information is obtained from the latter convolutional layer. A novel compound loss function was further employed to effectively differentiate the structure boundaries with significant HU variations and to retain the sharpness of the sCT image. The use of a mean squared distance (MSD) loss function in the general networks tends to produce images with blurry regions (Michael Mathieu, 2015). The generator loss function in this study consists of two losses: one is the adversarial loss ( $L_{adv}$ ) for distinguishing real images from synthetic images; the other is the distance loss ( $L_{distance}$ ) measured between real and synthetic images (Nie *et al.*, 2018) or between real and cycle images. The accuracy of the generator directly depends on design of the loss function. Suppose that the generator  $G$  obtains a synthetic image  $G(X) = Z$  from original image  $X$  to target image  $Y$ . A weighted summation of the two losses forms the compound loss function for the proposed method:

$$G = \arg \min_G \{ \lambda_{adv} L_{adv}(Z) + \lambda_{distance} L_{distance}(Z, Y) \}$$

Where  $\lambda_{adv}$  and  $\lambda_{distance}$  are balancing parameter. The adversarial loss function is defined as

$$L_{adv}(Z) = \text{MAD}(D(Z), 1)$$

in cycle GAN-based method (Zhu *et al.*, 2017). For distance loss  $L_{distance}(Z, Y)$ , we introduced a  $l_p$ -norm ( $p = 1.5$ ) distance, termed mean  $p$  distance (MPD). We also integrated an image gradient descent (GD) loss term into the loss function, with the aim of minimizing the difference of the magnitude of the gradient between the synthetic image and the original planning CT. In this way, the sCT will try to keep zones with strong gradients, such as edges, effectively compensating for the distance loss term. The generators are optimized as follows:

$$(G_{CT-MR}, G_{MR-CT})$$

$$= \arg \min_{G_{CT-MR}, G_{MR-CT}}$$

$$\left( \begin{aligned} & \lambda_{adv} \text{MAD}(D_{MR}(G_{CT-MR}(I_{CT})), 1) + \lambda_{MPD}^{cycle} \|G_{MR-CT}(G_{CT-MR}(I_{CT})), I_{CT}\|_p^p \\ & + \lambda_{GDL}^{cycle} \text{GDL}(G_{MR-CT}(G_{CT-MR}(I_{CT})), I_{CT}) \\ & + \lambda_{MPD}^{fake} \|G_{CT-MR}(I_{CT}), I_{MR}\|_p^p + \lambda_{GDL}^{fake} \text{GDL}(G_{CT-MR}(I_{CT}), I_{MR}) \\ & + \lambda_{adv} \text{MAD}(D_{CT}(G_{MR-CT}(I_{MR})), 1) + \lambda_{MPD}^{cycle} \|G_{CT-MR}(G_{MR-CT}(I_{MR})), I_{MR}\|_p^p \\ & + \lambda_{GDL}^{cycle} \text{GDL}(G_{CT-MR}(G_{MR-CT}(I_{MR})), I_{MR}) \\ & + \lambda_{MPD}^{fake} \|G_{MR-CT}(I_{MR}), I_{CT}\|_p^p + \lambda_{GDL}^{fake} \text{GDL}(G_{MR-CT}(I_{MR}), I_{CT}) \end{aligned} \right)$$

$$\text{GDL}(Z, Y) = \sum_{i,j,k}$$

$$\left( \begin{aligned} & \left\| \left| Z_{i,j,k} - Z_{i-1,j,k} \right| - \left| Y_{i,j,k} - Y_{i-1,j,k} \right| \right\|_2^2 + \left\| \left| Z_{i,j,k} - Z_{i,j-1,k} \right| - \left| Y_{i,j,k} - Y_{i,j-1,k} \right| \right\|_2^2 \\ & + \left\| \left| Z_{i,j,k} - Z_{i,j,k-1} \right| - \left| Y_{i,j,k} - Y_{i,j,k-1} \right| \right\|_2^2 \end{aligned} \right)$$

where  $\left\| \cdot \right\|_p$  denotes the  $L_p$ -norm, and  $\text{GDL}(\cdot)$  denotes the gradient descent loss function (Nie *et al.*, 2018).  $\lambda_{MPD}^{cycle}, \lambda_{GDL}^{cycle}, \lambda_{MPD}^{fake}, \lambda_{GDL}^{fake}, \lambda_{MPD}^{cycle}, \lambda_{GDL}^{cycle}, \lambda_{MPD}^{fake}, \lambda_{GDL}^{fake}$  are regularization parameters

for different regularization. The discriminator loss is computed by mean absolute distance (MAD) between the discriminator results of input synthetic and real images. To update all the hidden layers' kernels, the Adam gradient descent method was applied to minimize both generator loss and discriminator loss. Figure 1 outlines the workflow schematic of the proposed model, which consists of training and synthesizing stages. The training stage consists of 4 generators and 2 discriminators. Each generator includes several dense blocks. In the synthesizing stage, a new MR image is fed into the well-trained model to produce the sCT image.

The learning rate for Adam optimizer was set to  $2e^{-4}$ , and the model was trained and tested on an NVIDIA TITAN XP GPU with 12 GB of memory with a batch size of 8. During training, 3.4 GB CPU memory and 10.2 GB GPU memory was used for each batch optimization. The training was stopped after 150000 iterations. Training the model took around 15 hours, and sCT generation for one test patient took about 2 minutes.

## 2.D. Evaluation strategies

**2.D.1 Image quality**—To quantify the prediction quality, 3 commonly used metrics were applied, including mean absolute error (MAE), peak signal-to-noise ratio (PSNR), and normalized cross correlation (NCC) (Lei *et al.*, 2019b). MAE represents the discrepancies between the predictions and the reference HU numbers. PSNR is the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. NCC is a measure of similarity between 2 series as a function of the displacement of one relative to the other. The whole body MAE, PSNR and NCC were calculated together with the MAEs in soft tissue, bone and air pocket.

To evaluate the geometric displacement of the body contour and the bone contour between CT and sCT images, we calculated the Hausdorff distance 95% (HD95), mean surface distance (MSD), and the residual mean square distance (RMSD). The HD, MSD and RMSD metrics are generally used for quantification of boundary similarity between two surfaces. A displacement is associated with low HD, MSD and RMSD scores. The body surface was defined as the area with  $HU > -500$ . We first dilated the images to smooth the boundaries of the images, and then the air inside the images was removed by hole filling, and then we eroded the images to shrink the boundary to the original size. For the bone area,  $HU > 300$  was set as the threshold. Dilation was performed, followed by erosion to get a smooth boundary.

**2.D.1 Dosimetric analysis**—This patient cohort was previously treated with liver photon SBRT. The original CT images were transferred to RayStation (RaySearch Laboratories, Stockholm, Sweden) TPS (version 8A) for the proton treatment planning. Pencil beam scanning was used as the treatment technique. Experienced dosimetrists performed the planning and Monte Carlo (v4.2) was chosen as the dose calculation engine. OAR constraints were based on QUANTEC (Marks *et al.*, 2010) for conventional fraction of 1.8–2.0 Gy per fraction. One patient (p04) was decided to be excluded from this study because the patient has a large PTV contains not only tumor in liver but also spine and bowel lesions. This case is generally not suitable for proton therapy. For the rest of 20 patients, all

plans were prescribed with a total dose of 45 Gy in 25 fractions and normalized to 98% of the PTV receiving 100% of the prescription dose. For all the plans, two beams with optimized gantry angles were applied. The beam angles were chosen to minimize the normal tissues that involved in the beam path, to minimize the impact from patients' breathing motion and daily setup variation, and to avoid passing through heterogeneous tissues, like bone and bowel filled with air. Our proton beam has minimum energy of 70 MeV with Bragg peak at 4-cm water equivalent thickness (WET). If a tumor's depth at its proximal surface is less than 4 cm (WET), a range shifter with water equivalent depth of 2 cm, 3 cm, or 5 cm may be used to pull the proton ranged towards the surface.

After calculating the planning doses based on the original CT images, the evaluation doses calculation can be performed on the sCT images with the same beam settings. Given the purpose of acquiring MR images in this patient cohort was to help target volume delineation in the liver, the derived sCT images only included the area directly adjacent to the liver. The sCT thus have fewer axial slices than broader CTs that are necessary for conventional treatment planning. To make the OAR comparison feasible, original CT slices were added to the sCT to create data set of equal size. Since only coplanar beams were used, all the dose always fell into the area that was fully covered by the generated sCT. The dosimetric impact from the shared information between the two evaluation sources was zero.

In this study, the differences between the sCT and CT in PTV  $D_{10}$ ,  $D_{50}$ ,  $D_{95}$ ,  $D_{\text{mean}}$  and  $D_{\text{max}}$  were evaluated. Since liver proton therapy does not have beams passing through OARs, only liver and bowel that are close to target are considered. OAR  $D_{\text{mean}}$ ,  $D_{\text{max}}$ , and  $D_{10}$  together with some other clinical concerned dose-volume histogram (DVH) matrix including  $V_{15}$  and  $V_{20}$  were considered (Zeng *et al.*, 2017).

To evaluate the plane dose inaccuracy of the CT and sCT, the dose DICOM files were exported from RayStation to 3DVH (Sun Nuclear Corporation, Melbourne, USA). 3D global gamma analysis with 1%/1mm, 2%/2mm and 3%/3mm criteria with 10% dose threshold was carried out.

**2.D.2 Distal range analysis**—The planning CT- and sCT-based proton spread-out Bragg peak (SOBP) ranges of each individual beam along the beam-line that across the isocenter were retrieved from RayStation treatment planning system using an IronPython script. In this study, the proton beam range was defined at the 80% of the SOBP plateau dose at the distal range. The range difference and relative range difference between planning CT and sCT were calculated by:

$$\text{Range difference} = R_{80\text{-sCT}} - R_{80\text{-CT}}$$

$$\text{Relative range difference} = (R_{80\text{-sCT}} - R_{80\text{-CT}}) / R_{80\text{-CT}} \times 100\%$$

and compared with the Harvard Massachusetts General Hospital (MGH) uncertainty criteria:



$$\text{Uncertainty} < 3.5\% R_{80-CT} + 1 \text{ mm}$$

This criteria is used by MGH in proton treatment planning considering the uncertainty from organ motion, setup and anatomical variations, dose calculation approximations and biological considerations (Paganetti, 2012). Other institutions such as MD Anderson and University of Pennsylvania apply looser criteria (3.5%+3mm), while University of Florida has a tighter one (2.5%+1.5mm). Since the abdominal sCT prediction accuracy depends on the amplitude of organ motion, anatomical variations as well as the HU fidelity, the acceptable level of range difference can be guided by these uncertainty criteria.

### 3. RESULTS

#### 3.A. Image quality

Figure 2 lists the MAE, PSNR and NCC for each patient of this cohort. As summarized in Table 1, the mean ( $\pm$ standard deviation, abbreviation: SD) MAE, PSNR AND NCC are  $72.48 \pm 18.16$  HU,  $22.43 \pm 3.63$  dB, and  $0.92 \pm 0.04$  respectively.

The mean MAEs in bone, soft tissue, air pockets were  $216.81 \pm 63.05$  HU,  $58.62 \pm 30.61$  HU, and  $108.06 \pm 49.45$  HU, respectively.

Figure 3 shows MR and CT images and the sCT images of a representative patient. It is noticeable that the quality of the training MR images was not very good even after intensity inhomogeneity correction. The method is capable of handling MR intensity inhomogeneity, at least when the inhomogeneity effect is not significant, by producing relatively uniform HU numbers in the same tissue. Gentle motion artifacts can still be observed after deformable registration. Nonetheless, our deep learning-based method has shown promising results that with small HU difference and similar HU profile across regions with rapid HU change.

#### 3.B Dose comparison

Figure 4 exhibits the dose difference of two exemplary patients. The voxel dose differences were generally much less than 5% except at the distal edge of the beams. High dose discrepancy further occurs at the tissue and air interface. These results supported that proton therapy dose calculation is sensitive to the HU accuracy. Relatively large dose inaccuracies can be found with the presence of small rib bones and lung cavity. As can be seen in HU comparison graphs, the HU values of the livers are pretty close but discrepancies can be observed at the rib bones. This discrepancy can be due to the HU prediction inaccuracy or the rib bone displacement in CT and MR images because of patient motion. Overall however, our sCT has shown very promising results.

Mean gamma analysis pass rate of 1mm/1%, 2mm/2%, 3mm/3% criteria with 10% dose threshold were  $90.76 \pm 5.94\%$ ,  $96.98 \pm 2.93\%$  and  $99.37 \pm 0.99\%$ , respectively. Figure 5 shows the boxplot of the gamma analysis with different criteria.

Figure 6 shows the box plot of dose-volume statistics of PTV, liver, and bowel. The data in PTV and liver included the cohort of 20 patients, while limited patient data were included in the bowel (5 patients) because proton therapy has a very confined area of dose delivery and most of the patients had negligible dose deposition in OARs. As shown in the figure, one patient (p08) has relatively large dose differences in PTV Dmax and D95, otherwise the PTV dose-volume matrixes are all less than 0.5 Gy. In comparison to the prescribed dose of 45 Gy, the clinical impact of 0.5 Gy is insignificant (around 1%). As for the outlier, the Dmax difference of 1.5 Gy accounts for 3% dose difference, which is supposed to be clinically acceptable. The DVH differences in the liver were higher than those in PTV since HU prediction inaccuracy in the tissues such as bone and liver. The HU inaccuracies obviously affected the DVH differences in liver D10, V15 and V20 as shown in the figure. The maximum volume differences in liver V15 and V20 were both around 12.5 cm<sup>3</sup>. This value accounts for less than 1% of a typical liver volume of 1500 cm<sup>3</sup> (in this study, the liver volumes range from 1130 to 3021 cm<sup>3</sup> with an average value of 1709±593 cm<sup>3</sup>). In addition, due to the patient motion, organ positions such as those of the ribs were not the same between CT and MR images. The sCT organ locations are the same as the MR images, thus leading to a beam overshooting when the proton beam directly passes across the rib in the CT-based plan but not in the sCT-based plan. We consider it as one of the limitations of the ground truth employed in this study (the MR/CT pairs), but not the drawback of the proposed networks. Lastly, for the bowel DVHs, the differences are generally much smaller than 2.5 Gy. Similarly, the differences could be from the limitation of the ground truth because of the different bowel movement status during scans. Nonetheless, it highlights the importance of MRI-guided radiotherapy for dose delivery accuracy enhancement.

### 3.C Range evaluation

Figure 7 shows the proton beam range comparison between the plans created based on original CT and sCT. The range was retrieved from the dose grids in the beam-line direction that passes through the isocenter. It was more likely to reveal the maximum range difference because the pencil beams pass through the isocenter usually has the longest range. The largest absolute range difference and relative range difference was found in patient p08 (0.56 cm, 5.68 %) with maximum proton energy of 121 MeV. The median and mean absolute range differences were 0.17 and 0.186±0.155 cm, and the median and mean absolute relative range differences were 1.31 and 1.56±1.34 %. Using the Harvard MGH range uncertainty criteria shown in Figure 7(b), all beam ranges were within the tolerance level except two outliers (p05, p08).

## 4. DISCUSSION

This work sought to establish a novel method on generating liver sCT from corresponding MRI dataset by applying a dense-block cycle GAN model. To quantitatively evaluate the quality of the sCT, imaging endpoints (MAE, PSNR and NCC), proton treatment plan dosimetric endpoints (absolute dose difference, gamma analysis, and dose-volume statistics) and range endpoints (range difference, relative range difference, and Harvard MGH range uncertainty criteria) were performed. Side-by-side imaging comparisons revealed good agreement. The overall average MAE, PSNR and NCC of the sCT were 72.87±18.16 (HU),

22.65±3.63 dB and 0.92±0.04, respectively. These are competitive compared to counterpart values published from recent deep learning studies of sites such as brain and pelvis (Han, 2017; Emami *et al.*, 2018). The brain sCT generation based on GAN by Kazemifar *et al.* (Kazemifar *et al.*, 2019) achieved very good MAE with an average value of  $47.2 \pm 11.0$  HU. Photon VMAT plans were performed in their study and less than 1% dose differences were found for all of the DVH matrix. Some bone-air misclassification can still be seen in the sCT images and the tissue boundaries are relatively blurry. The cycle-GAN and compound loss used in our study can deal with those issues as better tissue boundaries and bone-air classification can be observed in our sCT images.

Although patient p05 and p08 had relatively higher MAE values and were later found with considerable range differences between plans created on CT and sCT, the interpretation of dosimetric and beam range accuracy in terms of these imaging endpoints is still limited, especially when the PSNR and NCC values does not appear to correspond to the dosimetric outcomes. It might be partly due to the sharp distal fall-off of the proton Bragg peak that confines the total area irradiated. Therefore, the overall imaging endpoint does not reveal very well the local mismatches or local HU difference that are important for proton treatment planning. In the application of liver proton therapy, since most of the beams must pass across the small rib bone before reaching the target, the accurate HU value prediction of the rib bone is very important. However, due to the small size, patient motion and general difficulty in bone prediction in sCT generation, the accurate rib bone prediction is particularly challenging. There are two published methods of abdomen sCT generation: one is based on fuzzy C-means and is not able to predict this small rib bone (Bredfeldt *et al.*, 2017); the other publication relies on atlas-based segmentation followed by voxel-based MR intensity to HU conversion based on predetermined conversion curves (Guerreiro *et al.*, 2019). The rib prediction of the latter depends on the position of atlas images, which might be totally different from the testing subject. As shown in Figure 4, our machine learning-based method was able to generate the rib one, but the local mismatch caused by patient motion, the accuracy of its HU number prediction is limited by showing dose discrepancies right in the direction that the proton passing across the rib bones. Besides HU inaccuracies, the rib displacement between the images obtained from MR and CT scans contributes to the dose differences. This is not the limitation of our deep-learning network, but the limitation of the ground truth used (imperfectly MR/CT pairs). It is important to note that cycle-GAN deals with the mismatches during the training stage to effectively and accurately learn a mapping between the intensities in CT and MR even when the two images are not perfectly aligned. However, during the synthesise stage, the prediction is solely depended on the MR images and the mapping algorithm developed during training. Therefore, the geometry of the sCT would be the same as the MR images. The mapping algorithm doesn't force the sCT structure geometry to be similar to the CT, which actually can benefit the future development of MRI-guided radiotherapy because we the organ positions may be different in real-time images as compared to the CT images. Figure 4 further revealed large dose differences caused by the local mismatch at the tissue and air interface. Because of the different physics interaction between proton and photon, the accuracy of proton dose calculation is much more sensitive to the HU values at the interfaces of different density tissues such as tissue-lung, tissue-bone, and bone edges.

Mean gamma analysis pass rate of 1mm/1%, 2mm/2%, 3mm/3% criteria with 10% dose threshold were  $90.76 \pm 5.94\%$ ,  $96.98 \pm 2.93\%$  and  $99.37 \pm 0.99\%$ , respectively. The results were comparable to the pelvis study done by Maspero *et al.* (Maspero et al., 2017) with a 98.4% pass rate of 2mm/2% criteria, and the brain and prostate study done by Klivula *et al.* (Klivila et al., 2016) with a 91% pass rate of 1mm/1% criteria. The statistics of OARs has shown significant dependence on the accuracy of the beam range accuracy.

Range evaluation was performed by retrieving the line dose along the beam-line direction that across the isocenter. The median absolute range difference was 0.17 cm with maximum value to be 0.56 cm. These values are higher than the data reported by Pileggi *et al.* brain study (Pileggi et al., 2018) with a median value of 0.05 cm and maximum of 0.44 cm and the Maspero *et al.* pelvis study (Maspero et al., 2017) with an average median of 0.01 cm. We believe that the main reason for our larger range displacement was due to the organ motion that caused mismatch between CT and sCT. In addition, different methods were used to retrieve the SOBP along the beam line direction: in this study it was based on the grid dose line that across the isocenter, which was more likely to reveal the maximum range. This work also reported the absolute median range difference which would be generally higher than the median value adopted by the other two studies. Overall, most of this study's range displacement fell into the MGH range uncertainty criteria acceptance level except two individual beams (among 20 patients multiplied by 2 beams each patient, a total of 40 beams). In one case (p05), it was due to the bowel movement. The other case (p08) was due to significant organ motion that blurred the sCT images that lead to failure to predict the rib bones.

As have discussed above, the imperfect image registration and patient motion contributed greatly to the discrepancies. Deformable image registration to the abdomen is still an open problem. Research in this area continues, but no practical solution to this problem has yet been found. The registered MR images after deformable alignment were blurred and distorted, depending on the quality of the original MR images. It affects the model training process that results in blurred and locally distorted sCT images. Unlike photon volumetric modulated arc therapy (VMAT) that employs multiple entry points that ultimately minimized the impact from local blur or distortion (Wang et al., 2019; Wang et al., 2018; Shafai-Erfani et al., 2019), proton dose calculation is very sensitive to this mismatch effect and exhibits noticeable disagreement in dose and range calculation. The dense block cycle GAN algorithm used in this work has, at least partly, lessened or avoided the local mismatch resulted from non-ideal registration, but the solution to ultimately resolve the problem is currently unavailable. The development of deformable image registration algorithm and motion management techniques therefore becomes fundamentally important to ensure the high quality of the machine learning training dataset and the outcome of predicted sCT images. Nonetheless, our algorithm has demonstrated its rigor under the current non-ideal registration conditions by showing comparable results with those sCT images generated at more stationary body sites such as brain and pelvis. The dosimetric and range agreement clearly warrants the further development of MR-based liver treatment planning.

A common issue inherent to MRI-only treatment planning is the MR image distortion. Its effect to highly conformal treatment including liver proton therapy can be serious (Seibert et

*et al.*, 2016; Wang *et al.*, 2013). At present, although not yet included as part of a standard package, the solution to correct such distortion have been supplied by many manufacturers (Jovicich *et al.*, 2006; Doran *et al.*, 2005; Baldwin *et al.*, 2007) and a standard guideline is under development (American Association of Physicists in Medicine Task Group No. 117). Together with the increased availability of commercially MRI simulators (Devic, 2012) and development of novel MRI image guidance (Oborn *et al.*, 2017b), high precision in target definition is the future for proton therapy. As we all know, very precise knowledge of the target delineation will do little good if we have incorrectly aligned the target at the treatment site. To avoid the large source of uncertainty from CT-MR registration, clearly, the unprecedented proton dose conformity calls for the more advanced MRI-only treatment process.

## 5. CONCLUSION AND FUTURE DIRECTIONS

We applied a novel learning-based approach to integrate dense-block into cycle GAN to synthesize abdominal sCT images from routine MR images for potential MRI-only liver proton therapy. The proposed method demonstrated a comparable level of precision in reliably generating sCT images for dose calculation, which supports further development of MRI-only treatment planning. Unlike photon therapy, the accuracy of proton dose calculation is highly dependent on stopping power rather than HU values. Therefore, the future directions of MR-only proton treatment planning include prediction of the stopping power map based on the MR images or generating elemental concentration maps that can be used for Monte Carlo simulations.

## ACKNOWLEDGMENTS

This research is supported in part by the National Cancer Institute of the National Institutes of Health under Award Number R01CA215718 (Yang) and R01CA184173 (Ren), and Emory Winship Cancer Institute pilot grant.

## REFERENCE

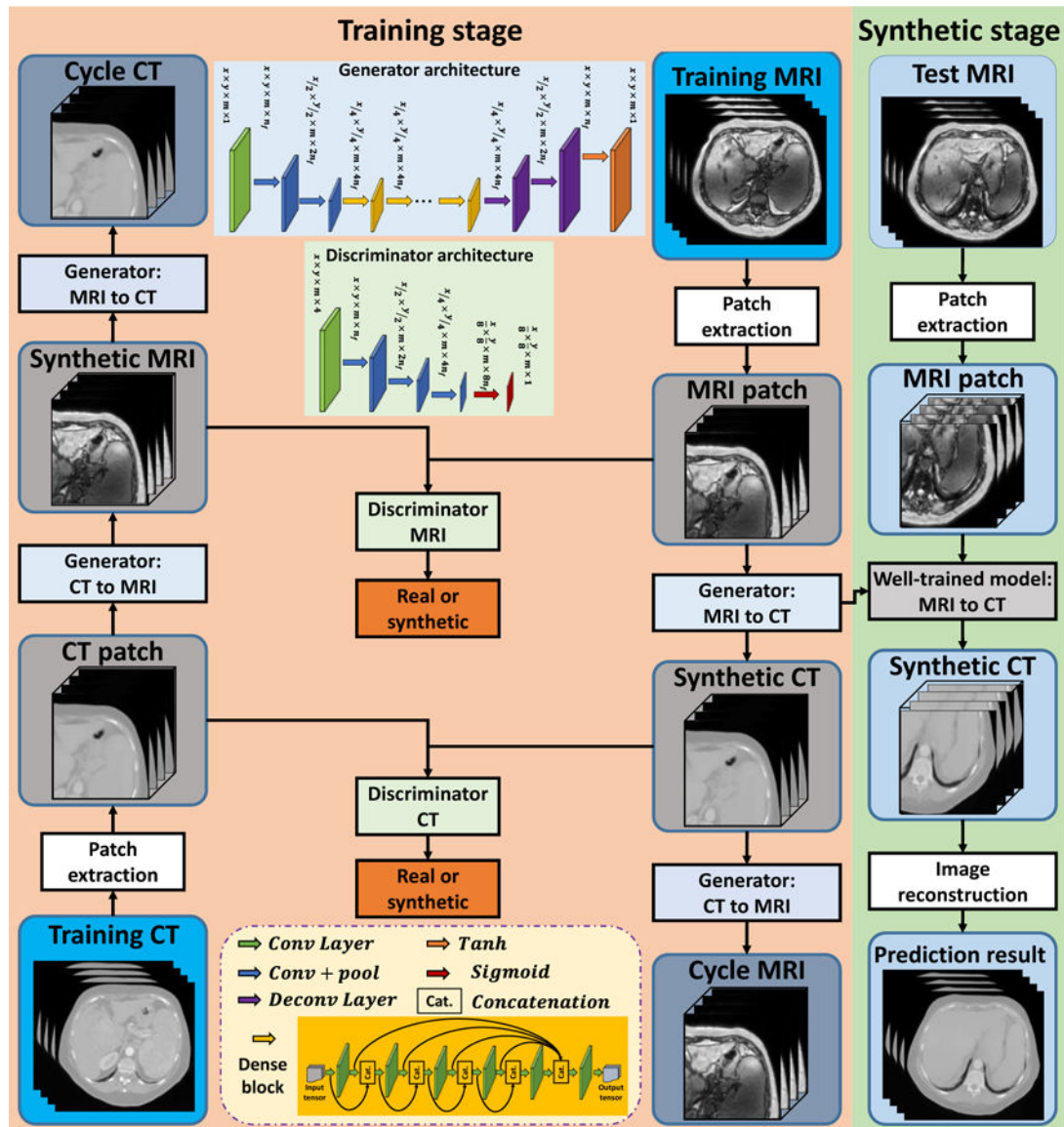
- Baldwin LN, Wachowicz K, Thomas SD, Rivest R and Fallone BG 2007 Characterization, prediction, and correction of geometric distortion in MR images *Med Phys* 34 388–99 [PubMed: 17388155]
- Bredfeldt JS, Liu L, Feng M, Cao Y and Balter JM 2017 Synthetic CT for MRI-based liver stereotactic body radiotherapy treatment planning *Phys Med Biol* 62 2922–34 [PubMed: 28306547]
- Chin AL, Lin A, Anamalayil S and Teo BKK 2014 Feasibility and limitations of bulk density assignment in MRI for head and neck IMRT treatment planning *J Appl Clin Med Phys* 15 100–11
- Daisne J-F, Sibomana M, Bol A, Cosnard G, Lonneux M, Grégoire VJR and Oncology 2003 Evaluation of a multimodality image (CT, MRI and PET) coregistration procedure on phantom and head and neck cancer patients: accuracy, reproducibility and consistency 69 237–45
- De Ruysscher D, Sterpin E, Haustermans K and Depuydt TJC 2015 Tumour movement in proton therapy: solutions and remaining questions: a review 7 1143–53
- Dean C, Sykes J, Cooper R, Hatfield P, Carey B, Swift S, Bacon S, Thwaites D, Sebag-Montefiore D and Morgan AJTB 2012 An evaluation of four CT–MRI co-registration techniques for radiotherapy treatment planning of prone rectal cancer patients 85 61–8
- Delaney G, Jacob S, Featherstone C and Barton M 2005 The role of radiotherapy in cancer treatment 104 1129–37
- Devic S 2012 MRI simulation for radiotherapy treatment planning *Med Phys* 39 6701–11 [PubMed: 23127064]

- Doran SJ, Charles-Edwards L, Reinsberg SA and Leach MO 2005 A complete distortion correction for MR images: I. Gradient warp correction *Phys Med Biol* 50 1343 [PubMed: 15798328]
- Dougeni E, Faulkner K and Panayiotakis G 2012 A review of patient dose and optimisation methods in adult and paediatric CT scanning *European Journal of Radiology* 81 e665–e83 [PubMed: 21684099]
- Edmund JM and Nyholm T 2017 A review of substitute CT generation for MRI-only radiation therapy *Radiation oncology (London, England)* 12 28-
- Emami H, Dong M, Nejad-Davarani SP and Glide-Hurst CK 2018 Generating synthetic CTs from magnetic resonance images using generative adversarial networks *Medical Physics* 45 3627–36
- Guerreiro F, Burgos N, Dunlop A, Wong K, Petkar I, Nutting C, Harrington K, Bhide S, Newbold K, Dearnaley D, deSouza NM, Morgan VA, McClelland J, Nill S, Cardoso MJ, Ourselin S, Oelfke U and Knopf AC 2017 Evaluation of a multi-atlas CT synthesis approach for MRI-only radiotherapy treatment planning *Med Phys* 35 7–17
- Guerreiro F, Koivula L, Seravalli E, Janssens GO, Maduro JH, Brouwer CL, Korevaar EW, Knopf AC, Korhonen J and Raaymakers BW 2019 Feasibility of MRI-only photon and proton dose calculations for pediatric patients with abdominal tumors *Physics in Medicine & Biology* 64 055010 [PubMed: 30669135]
- Han X 2017 MR-based synthetic CT generation using a deep convolutional neural network method *Med Phys* 44 1408–19 [PubMed: 28192624]
- Hsu S-H, Cao Y, Huang K, Feng M and Balter JM 2013 Investigation of a method for generating synthetic CT models from MRI scans of the head and neck for radiation therapy *Phys Med Biol* 58 8419 [PubMed: 24217183]
- Huynh T, Gao Y, Kang J, Wang L, Zhang P, Lian J, Shen D and Alzheimer's Disease Neuroimaging I 2016 Estimating CT Image From MRI Data Using Structured Random Forest and Auto-Context Model *IEEE transactions on medical imaging* 35 174–83 [PubMed: 26241970]
- Jovicich J, Czanner S, Greve D, Haley E, van Der Kouwe A, Gollub R, Kennedy D, Schmitt F, Brown G and MacFall J 2006 Reliability in multi-site structural MRI studies: effects of gradient non-linearity correction on phantom and human data *Neuroimage* 30 436–43 [PubMed: 16300968]
- Kanai T, Kawachi K, Kumamoto Y, Ogawa H, Yamada T, Matsuzawa H and Inada TJM p 1980 Spot scanning system for proton radiotherapy *7* 365–9
- Kazemifar S, McGuire S, Timmerman R, Wardak Z, Nguyen D, Park Y, Jiang S and Owrangi A 2019 MRI-only brain radiotherapy: Assessing the dosimetric accuracy of synthetic CT images generated using a deep learning approach *Radiotherapy and Oncology* 136 56–63 [PubMed: 31015130]
- Koivula L, Wee L and Korhonen JJM p 2016 Feasibility of MRI-only treatment planning for proton therapy in brain and prostate cancers: Dose calculation accuracy in substitute CT images *Med Phys* 43 4634–42 [PubMed: 27487880]
- Kontaxis C, Bol GH, Stemkens B, Glitzner M, Prins FM, Kerkmeijer LGW, Legendijk JJW and Raaymakers BW 2017 Towards fast online intrafraction replanning for free-breathing stereotactic body radiation therapy with the MR-linac *Physics in Medicine & Biology* 62 7233–48 [PubMed: 28749375]
- Korhonen J, Kapanen M, Keyriläinen J, Seppälä T and Tenhunen M 2014 A dual model HU conversion from MRI intensity values within and outside of bone segment for MRI-based radiotherapy treatment planning of prostate cancer *Med Phys* 41
- Legendijk JJW, Raaymakers BW, Van den Berg CAT, Moerland MA, Philippens ME and van Vulpen M 2014 MR guidance in radiotherapy *Physics in Medicine and Biology* 59 R349–R69 [PubMed: 25322150]
- Largent A, Barateau A, Nunes J-C, Lafond C, Greer PB, Dowling JA, Saint-Jalmes H, Acosta O and de Crevoisier R 2018 Pseudo-CT generation for MRI-only radiotherapy treatment planning: comparison between patch-based, atlas-based, and bulk density methods *Int J Radiat Oncol Biol Phys*
- Lei Y, Harms JM, Wang T, Tian S, Zhou J, Shu H-K, Zhong J, Mao H, Curran WJ and Liu T 2019a MRI-based synthetic CT generation using semantic random forest with iterative refinement *Physics in medicine and biology*

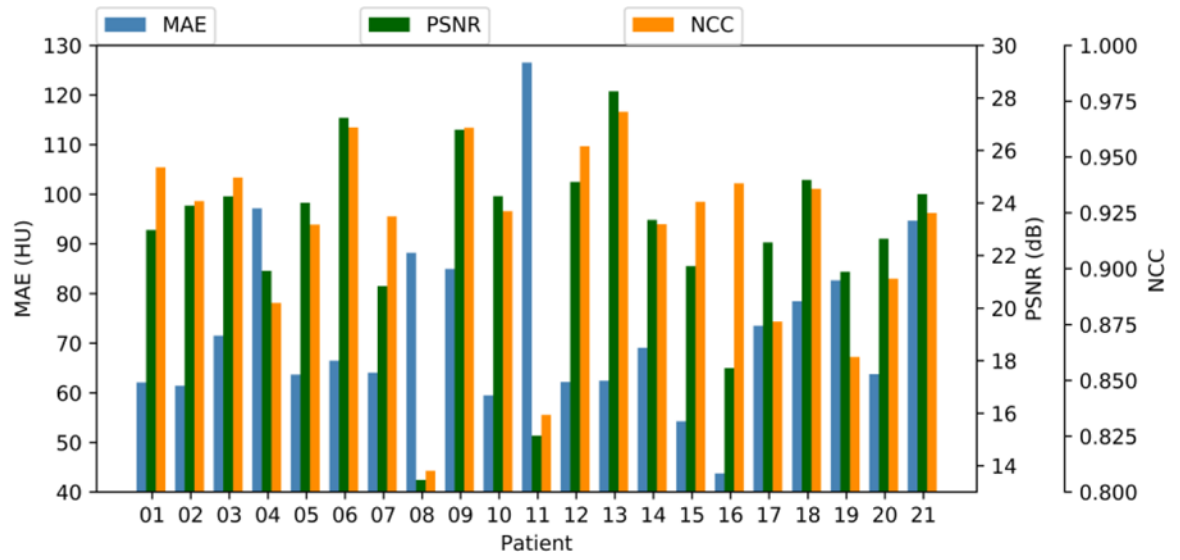
- Lei Y, Jeong JJ, Wang T, Shu H-K, Patel P, Tian S, Liu T, Shim H, Mao H and Jani AB 2018a MRI-based pseudo CT synthesis using anatomical signature and alternating random forest with iterative refinement model *Journal of Medical Imaging* 5 043504 [PubMed: 30840748]
- Lei Y, Shu H, Tian S, Wang T, Liu T, Mao H, Shim H, Curran W and Yang X 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2018b), vol. Series): IEEE) pp 5150–3
- Lei Y, Tang X, Higgins K, Lin J, Jeong J, Liu T, Dhabaan A, Wang T, Dong X, Press R, Curran WJ and Yang X 2019b Learning-based CBCT correction using alternating random forest based on auto-context model *Medical Physics* 46 601–18 [PubMed: 30471129]
- Lei Y, Wang T, Liu Y, Higgins K, Tian S, Liu T, Mao H, Shim H, Curran WJ, Shu H-K and Yang X 2019c) *Medical Imaging 2019: Physics of Medical Imaging*, vol. Series): International Society for Optics and Photonics)
- Levin W, Kooy H, Loeffler J and DeLaney T J B j o C 2005 Proton beam therapy 93 849
- Marks LB, Yorke ED, Jackson A, Ten Haken RK, Constine LS, Eisbruch A, Bentzen SM, Nam J and Deasy JO 2010 Use of normal tissue complication probability models in the clinic *International journal of radiation oncology, biology, physics* 76 S10–S9
- Maspero M, Van den Berg CA, Landry G, Belka C, Parodi K, Seevinck PR, Raaymakers BW and Kurz C 2017 Feasibility of MR-only proton dose calculations for prostate cancer radiotherapy using a commercial pseudo-CT generation method *Phys Med Biol* 62 9159 [PubMed: 29076458]
- Michael Mathieu CC, LeCun Yann 2015 Deep multi-scale video prediction beyond mean square error CoRR <http://arxiv.org/abs/1511.05440>
- Nie D, Trullo R, Lian J, Petitjean C, Ruan S, Wang Q and Shen D 2017 Medical Image Synthesis with Context-Aware Generative Adversarial Networks pp 417–25
- Nie D, Trullo R, Lian J, Wang L, Petitjean C, Ruan S, Wang Q and Shen D 2018 Medical Image Synthesis with Deep Convolutional Adversarial Networks *IEEE Trans Biomed Eng*
- Nyholm T, Nyberg M, Karlsson MG and Karlsson M J R o 2009 Systematisation of spatial uncertainties for comparison between a MR and a CT-based radiotherapy workflow for prostate treatments 4 54
- Oborn BM, Dowdell S, Metcalfe PE, Crozier S, Mohan R and Keall PJ 2017a Future of medical physics: Real-time MRI-guided proton therapy *Medical Physics* 44 e77–e90 [PubMed: 28547820]
- Oborn BM, Dowdell S, Metcalfe PE, Crozier S, Mohan R and Keall PJ 2017b Future of medical physics: Real-time MRI-guided proton therapy 44 e77–e90
- Paganetti H 2012 Range uncertainties in proton therapy and the role of Monte Carlo simulations *Physics in medicine and biology* 57 R99–R117 [PubMed: 22571913]
- Pileggi G, Speier C, Sharp GC, Izquierdo Garcia D, Catana C, Pursley J, Amato F, Seco J and Spadea MF 2018 Proton range shift analysis on brain pseudo-CT generated from T1 and T2 MR *Acta Oncologica* 57 1521–31 [PubMed: 29842815]
- Roberson PL, McLaughlin PW, Narayana V, Troyer S, Hixson GV and Kessler MLJM p 2005 Use and uncertainties of mutual information for computed tomography/magnetic resonance (CT/MR) registration post permanent implant of the prostate 32 473–82
- Seibert TM, White NS, Kim G-Y, Moiseenko V, McDonald CR, Farid N, Bartsch H, Kuperman J, Karunamuni R and Marshall D 2016 Distortion inherent to magnetic resonance imaging can lead to geometric miss in radiosurgery planning *Pract Radiat Oncol* 6 e319–e28 [PubMed: 27523440]
- Shafai-Erfani G, Wang T, Lei Y, Tian S, Patel P, Jani AB, Curran WJ, Liu T and Yang X 2019 Dose evaluation of MRI-based synthetic CT generated using a machine learning method for prostate cancer radiotherapy *Medical Dosimetry*
- Sjölund J, Forsberg D, Andersson M and Knutsson H 2015 Generating patient specific pseudo-CT of the head from MR using atlas-based regression *Phys Med Biol* 60 825 [PubMed: 25565133]
- Ulin K, Urie MM and Cherlow J M J I J o R O B P 2010 Results of a multi-institutional benchmark test for cranial CT/MR image registration 77 1584–9
- Wang H, Balter J and Cao Y 2013 Patient-induced susceptibility effect on geometric distortion of clinical brain MRI for radiation treatment planning on a 3T scanner *Phys Med Biol* 58 465 [PubMed: 23302471]

- Wang T, Lei Y, Manohar N, Tian S, Jani AB, Shu H-K, Higgins K, Dhabaan A, Patel P, Tang X, Liu T, Curran WJ and Yang X 2019 Dosimetric study on learning-based cone-beam CT correction in adaptive radiation therapy *Medical Dosimetry*
- Wang T, Manohar N, Lei Y, Dhabaan A, Shu H-K, Liu T, Curran WJ and Yang X 2018 MRI-based treatment planning for brain stereotactic radiosurgery: Dosimetric validation of a learning-based pseudo-CT generation method *Medical Dosimetry*
- Wen N, Guan H, Hammoud R, Pradhan D, Nurushev T, Li S and Movsas B 2007 Dose delivered from Varian's CBCT to patients receiving IMRT for prostate cancer *Physics in Medicine and Biology* 52 2267–76 [PubMed: 17404468]
- Wolterink JM, Dinkla AM, Savenije MH, Seevinck PR, van den Berg CA and Išgum I *International Workshop on Simulation and Synthesis in Medical Imaging, 2017*, vol. Series: Springer) pp 14–23
- Yang X, Wang T, Lei Y, Higgins K, Liu T, Shim H, Curran WJ, Mao H and Nye JA 2019 MRI-based attenuation correction for brain PET/MRI based on anatomic signature and machine learning *Physics in Medicine & Biology* 64 025001 [PubMed: 30524027]
- Zeng Z-C, Seong J, Yoon SM, Cheng JC-H, Lam K-O, Lee A-S, Law A, Zhang J-Y and Hu Y 2017 Consensus on stereotactic body radiation therapy for small-sized hepatocellular carcinoma at the 7th Asia-Pacific Primary Liver Cancer Expert Meeting *Liver cancer* 6 264–74 [PubMed: 29234630]
- Zhu JY, Park T, Isola P and Efros AA 2017 Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks *Ieee I Conf Comp Vis* 2242–51

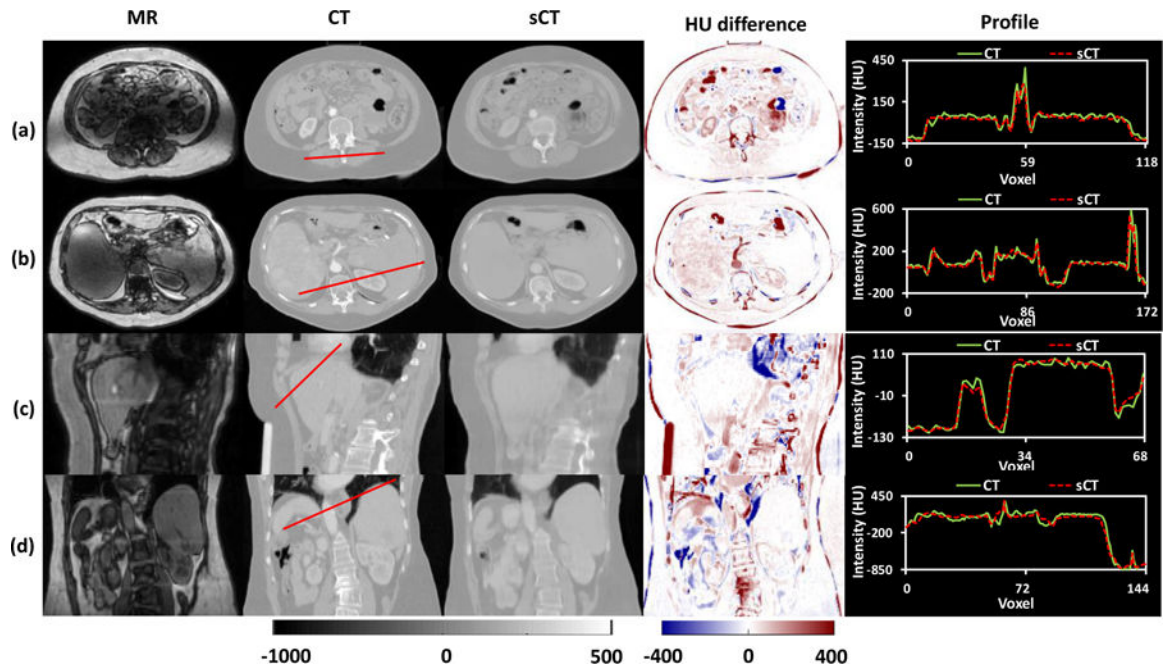




**Figure 1.** Schematic flow chart of the proposed algorithm for MRI-based sCT generation. The training stage is consisted of 4 generators and 2 discriminators. Each generator includes several dense blocks. The synthesizing stage is shown on the right side, in which a new MR image is fed into this well-trained model to produce the sCT.

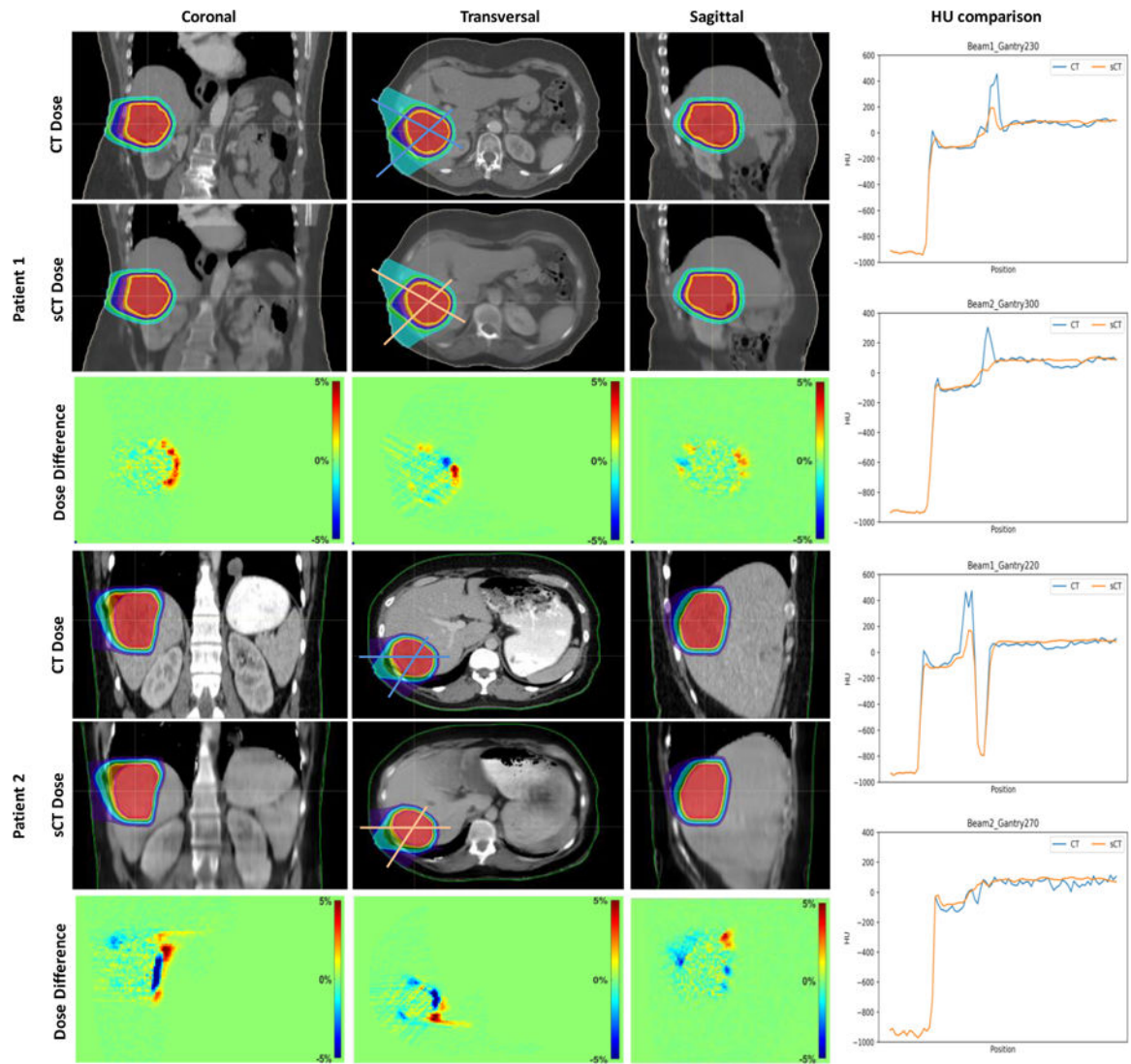


**Figure 2.**  
Results of the MAE, PSNR and NCC for each patient.

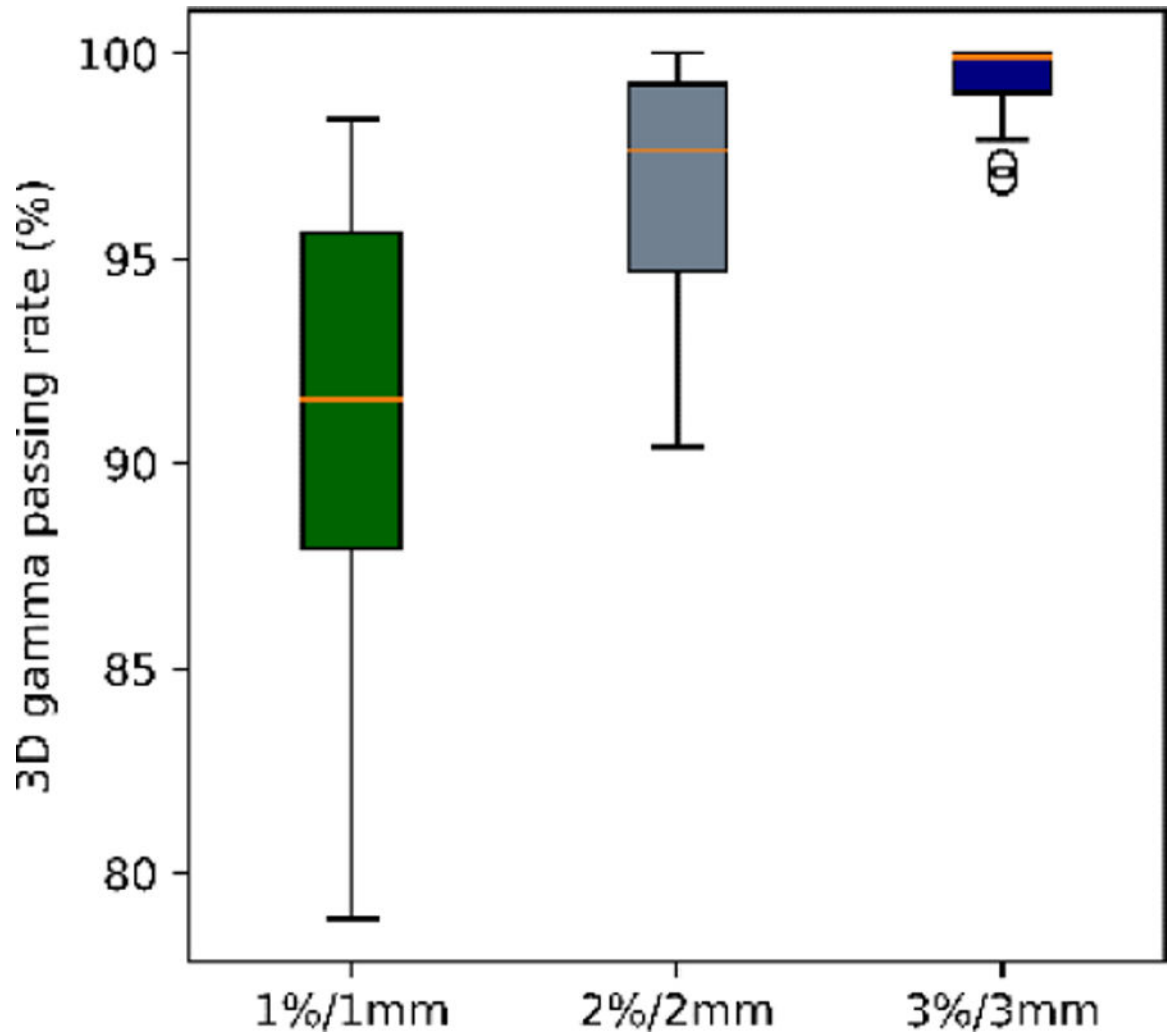


**Figure 3.**

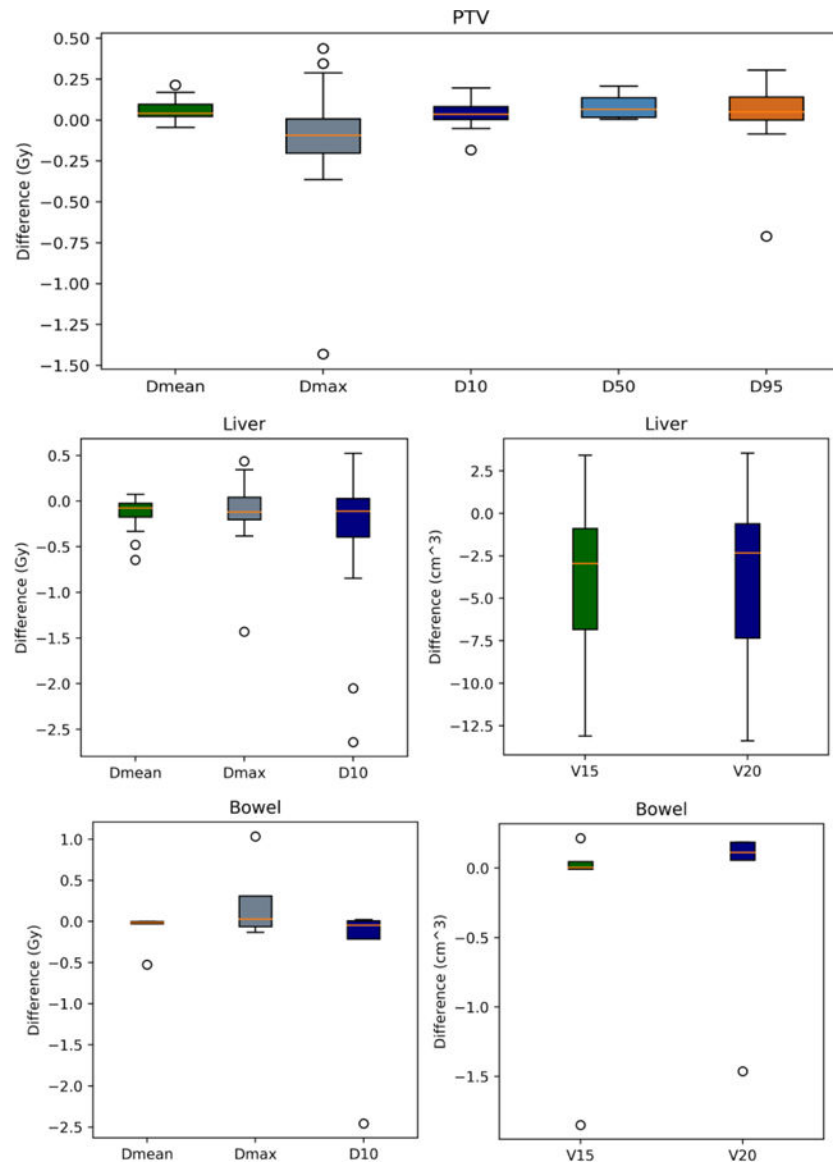
From left to right: MR image, CT image, sCT image, HU difference image between CT and sCT images, plot profile of red line in CT and sCT images. (a) and (b) Transversal view of a patient's abdominal images. (a) presents the site with a number of organs and vertebral bone. (b) presents the liver site that has small tiny rib bones. (c) Sagittal view. (d) Coronal view.



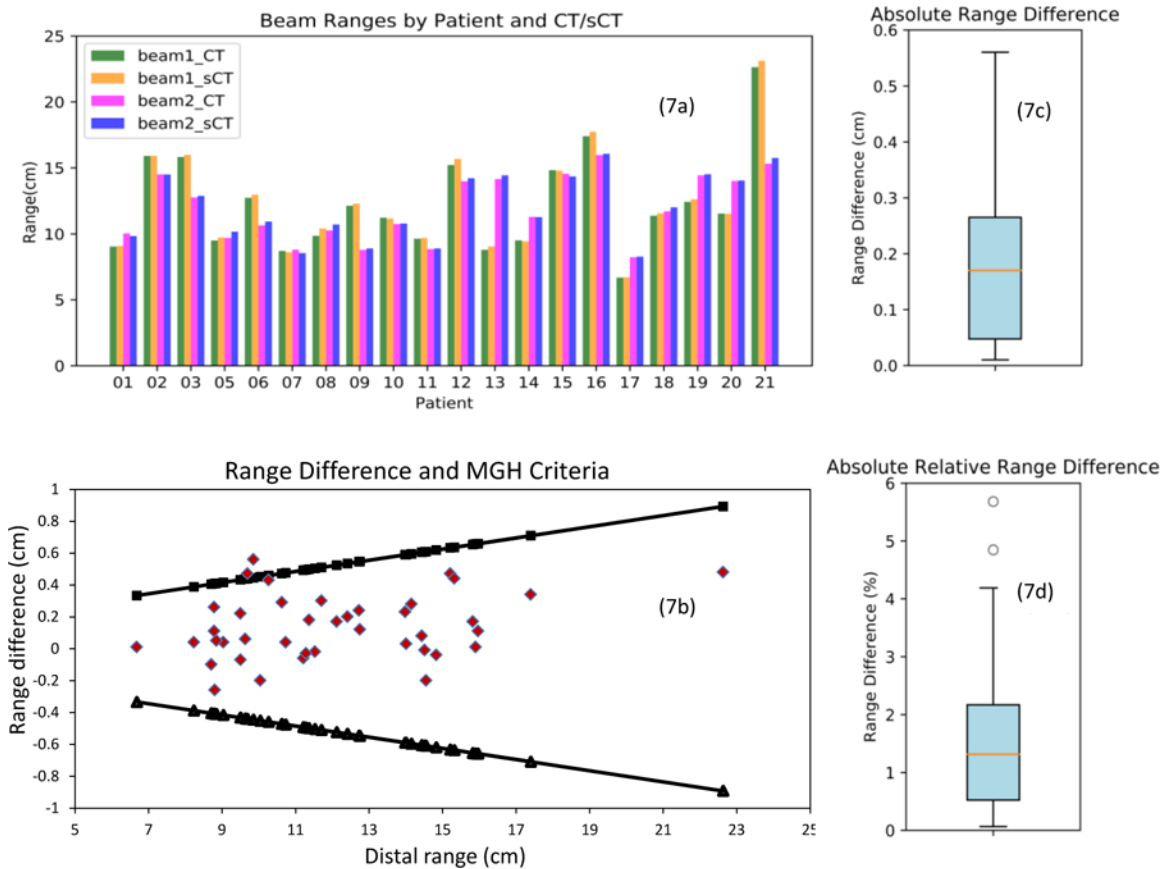
**Figure 4.** From left to right: coronal, transversal and sagittal view. 2 exemplary patients were used to demonstrate the dose differences between plans calculated on original CT and sCT. The dose profiles were retrieved from the 3 different views that interest with the isocenter.



**Figure 5.** Gamma passing rates for 3 criteria: 1mm/1%, 2mm/2%, and 3mm/3% with 10% dose threshold. The central orange line indicates the median value, and the borders of the box represent the 25<sup>th</sup> and 75<sup>th</sup> percentiles. The outliers are plotted by the black “O” marker.



**Figure 6.** Box plot of DVH difference between sCT and CT for the PTV and OARs. The central orange line indicates the median value, and the borders of the box represent the 25<sup>th</sup> and 75<sup>th</sup> percentiles. The outliers are plotted by the black “O” marker.



**Figure 7.** Range comparison between the plans created on CT and sCT. (7a) Beam ranges of each beam of the 20 patient cohort. (7b) The red rhombus marker shows the distribution of the range differences as a function of the actual range value from the plan calculated on the original CT. The black square and triangle markers and the black lines represent the upper and lower limit for the MGH range uncertainty criteria. (7c) Box plot of absolute range difference. (7d) Box plot of absolute relative range difference.

**Table 1.**

Statistics for the MAE, PSNR and NCC values of the cohort.

	<b>Mean (<math>\pm</math>SD)</b>	<b>Median</b>	<b>Min</b>	<b>Max</b>
MAE (HU)	72.87 $\pm$ 18.16	66.46	43.74	126.53
PSNR (dB)	22.65 $\pm$ 3.63	23.35	13.46	28.25
NCC	0.92 $\pm$ 0.04	0.93	0.81	0.97

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript