PAPER

Abdominal synthetic CT reconstruction with intensity projection prior for MRI-only adaptive radiotherapy

Sven Olberg1,2, Jaehee Chun3, Byong Su Choi1,3, Inkyung Park4, Hyun Kim4, Taeho Kim5, Jin Sung Kim5, Olga Green4 and Justin C Park1

1 Medical Artificial Intelligence and Automation (MAIA) Laboratory, Department of Radiation Oncology, University of Texas Southwestern Medical Center, Dallas, TX 75390, United States of America
2 Department of Biomedical Engineering, Washington University in St. Louis, St. Louis, MO 63110, United States of America
3 Department of Radiation Oncology, Yonsei Cancer Center, Yonsei University College of Medicine, Seoul, Republic of Korea
4 Department of Radiation Oncology, Washington University in St. Louis, St. Louis, MO 63110, United States of America

E-mail: jinsung@yuhs.ac

Keywords: MRI-guided ART, deep learning, synthetic CT

Abstract

Objective. Owing to the superior soft tissue contrast of MRI, MRI-guided adaptive radiotherapy (ART) is well-suited to managing interfractional changes in anatomy. An MRI-only workflow is desirable, but producing synthetic CT (sCT) data through paired data-driven deep learning (DL) for abdominal dose calculations remains a challenge due to the highly variable presence of intestinal gas. We present the preliminary dosimetric evaluation of our novel approach to sCT reconstruction that is well suited to handling intestinal gas in abdominal MRI-only ART.

Approach. We utilize a paired data DL approach enabled by the intensity projection prior, in which well-matching training pairs are created by propagating air from MRI to corresponding CT scans. Evaluations focus on two classes: patients with (1) little involvement of intestinal gas, and (2) notable differences in intestinal gas presence between corresponding scans. Comparisons between sCT-based plans and CT-based clinical plans for both classes are made at the first treatment fraction to highlight the dosimetric impact of the variable presence of intestinal gas.

Main results. Class 1 patients (n = 13) demonstrate differences in prescribed dose coverage of the PTV of 1.3 ± 2.1% between clinical plans and sCT-based plans. Mean DVH differences in all structures for Class 1 patients are found to be statistically insignificant. In Class 2 (n = 20), target coverage is 13.3 ± 11.0% higher in the clinical plans and mean DVH differences are found to be statistically significant. Significance. Significant deviations in calculated doses arising from the variable presence of intestinal gas in corresponding CT and MRI scans result in uncertainty in high-dose regions that may limit the effectiveness of adaptive dose escalation efforts. We have proposed a paired data-driven DL approach to sCT reconstruction for accurate dose calculations in abdominal ART enabled by the creation of a clinically unavailable training data set with well-matching representations of intestinal gas.

1. Introduction

The proliferation of magnetic resonance imaging (MRI)-based radiotherapy delivery systems in recent years has pushed applications of MRI-guided radiation therapy (MRgRT) to the forefront of RT (Wen et al., Fallone 2014, Keall et al. 2014, Mutic and Dempsey 2014, Fischer-Valuck et al. 2017, Pollard et al. 2017, Raaymakers et al. 2017). The superior soft tissue contrast of MRI compared to that of x-ray computed tomography (CT) improves target delineation in many sites, enabling near real-time motion tracking and management during treatment (Noel et al. 2015, Raaymakers et al. 2017). The advantages of MRI also lend themselves well to managing interfractional changes in a patient’s anatomy, as MRI setup scans acquired in an adaptive radiotherapy (ART) workflow capture the anatomy of the day without exposing the patient to additional ionizing radiation (Fischer-Valuck...
et al 2017). The need for electron density information in dose calculations, however, necessitates that these setup scans be registered to a CT simulation scan that may have been acquired weeks prior to a given treatment fraction (Schmidt and Payne 2015). Issues may arise in some cases when the anatomy represented in each scan is incompatible due to changes in the geometry and position of organs of interest, which is a challenge that is especially relevant to gastrointestinal (GI) structures and intestinal gas pockets considered during ART in the abdomen. In these cases, the ideal, unconstrained MRgRT workflow would involve intensive manual contouring so that density overrides could be performed in order to approximate in the CT simulation scan the anatomy of the day captured in the MRI setup scan.

It is this potential for challenge and uncertainty that comes with multi-modal image registration and intensive contouring that makes an MRI-only workflow—one in which MRI is the sole imaging modality used for planning and guidance—an attractive alternative to the conventional MRgRT workflow. The primary challenge in an MRI-only workflow is generating synthetic CT (sCT) data that yields the electron density information used in dose calculations. Many existing approaches to this task have been summarized as belonging to three general classes: atlas-based, voxel-based, and learning-based methods (Edmund and Nyholm 2017, Johnstone et al 2018). Recent investigations have focused primarily on approaches belonging to the last category, namely deep learning (DL)-based approaches in which convolutional neural networks are used to approximate a mapping between MRI and CT images (Spadea et al 2021).

Studies in this space have primarily been limited to regions of relatively static anatomy, including the head & neck (Han 2017) and general pelvis (Arabi et al 2018, Chen et al 2018, Maspero et al 2018). More recently, our group has extended these investigations into the thorax, where registration is challenged by pulmonary and cardiac motion, by generating sCT data for use in MRI-only breast RT (Olberg et al 2019). When considering applications in the more dynamic region of the abdomen, a primary challenge to employing these DL strategies becomes readily apparent: the majority of frameworks require paired data for training. The presence and hard-to-characterize motion of intestinal gas gives rise to observable differences in bowel filling and position on numerous time scales: seconds in the course of a single scan, minutes during treatment delivery, and hours between MRI and CT scans used for treatment planning (Nakamoto et al 2004, Feng et al 2009, Kumagai et al 2009, Mostafaei et al 2018, Corradini et al 2019). This challenge is even more relevant clinically when considering setup scans in MRI-guided adaptive treatments. In these situations, MRI setup scans acquired the day of treatment are registered to the CT simulation scan that may have been acquired weeks previously. When large discrepancies in bowel filling and position exist, the MRI setup scan and simulation CT scan may be rendered largely incompatible. Corrective measures that could be undertaken in an online adaptive workflow represent a potentially significant delay to treatment delivery while the patient remains on-table (Henke et al 2018). As an alternative to the paired data approaches, one may consider the application of an unpaired style-transfer approach exemplified by CycleGAN in settings in which abundant, unpaired data exists (Zhu et al 2018, Spadea et al 2021). However, the authors of the original CycleGAN paper acknowledge a gap between the paired and unpaired results that is hard or even impossible to close in some settings, especially those in which there exists some inherent ambiguity (Zhu et al 2018). This conclusion is mirrored in other studies of the sCT reconstruction task; (Peng et al 2020) conclude that the conditional paired approach was ‘preferable if high-quality MR–CT pairs were available’ in the nasopharynx, which is another area challenged by the presence of air, and (Fu et al 2020) demonstrate no benefit to adopting the CycleGAN in the upper abdomen for liver cancer patients.

In light of these challenges, many existing studies on sCT reconstruction in the abdomen have abandoned the DL-based approaches in favor of various classification or thresholding-based approaches wherein manual steps are taken in the process of generating sCT data to account for the presence of air. Bredfeldt et al (2017) and Hsu et al (2019) utilized a fuzzy c-means clustering algorithm to classify tissue classes based on multiple MRI volumes captured using different sequences, taking care to threshold image regions where air is expected to be found before applying the classification algorithm. Alternatively, (Ahunbay et al 2019) opted to use deformable image registration between the daily MRI and simulation CT scans to transfer electron density information while using thresholding operations in manually contoured regions to identify the presence of air. Guerreiro et al (2019) explored a hybrid atlas and intensity-based conversion algorithm (Korhonen et al 2014, Koivula et al 2016, 2017) in which contoured regions of air were transferred directly from the simulation CT after Hounsfield units (HU) were assigned. Most recently, both (Cusumano et al 2020) and (Qian et al 2020) have reported experiences with DL-based sCT reconstruction in the abdomen. Neither study, however, deals explicitly with the issue of air discrepancies between corresponding MRI and CT scans. Cusumano et al (2020) instead employed exclusion criteria to select only MR and CT images with ‘high correspondence in terms of air pockets location between the two images’ for training and excluded a further six patients from the test set for the reason of poor correspondence of air between the two modalities. Similarly, (Qian et al 2020) do not discuss results for patients with poorly matched representations of air.

In the present study, we return to paired data-driven DL with a novel hybrid approach to the abdominal sCT reconstruction task enabled by the creation of a training data set that is clinically unavailable. As was previously
discussed, the primary barrier to the adoption of many DL-based algorithms in this setting is the requirement for paired training data. Mismatches in the presence of intestinal gas between corresponding MRI and CT scans render the collection of a training data set of sufficient size an impossible task. Considering this and the challenges faced with an unpaired approach, we first utilize automated thresholding and morphological reconstruction operations to identify and propagate regions of air between corresponding MR and CT images to produce a well-matched training data set. We present here the preliminary evaluation of our paired data DL-based approach to sCT reconstruction in the abdomen enabled by the novel utilization of the intensity projection prior with a focus on showcasing the effects of intestinal gas differences on dose calculations. Dosimetric comparisons are made between two classes of test patients separated by a qualitative distinction made at the time of data collection prior to any evaluation: Class 1, consisting of well-matched patients demonstrating little involvement of intestinal gas, and Class 2, consisting of patients characterized by notable differences in the presence of intestinal gas in corresponding MRI and CT scans. Dosimetric accuracy is established using the patients of Class 1 while comparisons of target coverage between the sCT-based plans and the simulation CT-based clinical plans for patients of Class 2 highlight the complications posed by intestinal gas during MRI-only ART in the abdomen.

2. Materials and methods

2.1. Patient population
Data sets used in the present study were retrospectively collected from a population of pancreatic cancer patients previously treated at our institution using MRgRT. In each case, patients underwent CT and MRI simulation scans prior to treatment planning. Scans were acquired in treatment position using an Alpha Cradle (Smithers Medical Products Inc., North Canton, OH) with no additional immobilization devices. The nominal prescription in the selected population was a total dose of 50 Gy delivered to 95% of the PTV in 5 fractions. A small number of cases deviated from this prescription with total prescribed doses ranging from 30 to 40 Gy delivered in 5 fractions. For the nominal prescription, satisfying dose-volume constraints of $<0.5 \text{ cm}^3$ at 36 Gy for notable structures including the duodenum, stomach, small bowel, and large bowel was prioritized over target coverage. In any case in which one of these constraints was violated, the calculated dose was normalized to satisfy the violated constraint. The same normalization applied in the clinical case was also applied in the evaluation of our proposed method as discussed later. In all patient cases involved here, dose calculations in the clinical plan were performed using electron density data derived directly from the simulation CT scan with no additional density overrides included to account for the presence of air.

2.2. Training and testing data
A total of 89 patient data sets (one pair of corresponding MRI and CT scans per patient) were used in the present study, randomly assigned in the following splits: 53 train / 3 valid / 33 test. Ten-fold validation of the framework in the sCT reconstruction task has been previously carried out by our group, and an additional 3-holdout validation is performed here (Olberg et al 2019). The 33 test patients were qualitatively subdivided at the time of data collection prior to testing into the two classes mentioned previously: 13 well-matched patients in Class 1 and 20 patients characterized by notable discrepancies in bowel filling in Class 2. CT simulation scans acquired using a dedicated simulation machine (Brilliance CT, Philips Medical Systems, Andover, MA) were registered with 0.35 T MRI scans (nominally $276 \times 276 \times 80$ matrix, $1.63 \times 1.63 \times 3 \text{ mm}^3$) acquired using the MRIdian system (ViewRay Inc., Oakwood Village, OH) with a bSSFP sequence before being exported for pre-processing. Processing yielded a training data set of 2017 paired images, which were padded to dimensions of $520 \times 520$ before training via $320 \times 320$ random crops. The framework was trained for 1500 epochs with a batch size of 1 using TensorFlow (Abadi et al 2015) v1.7.0 in Python running on a 12 GB Titan Xp GPU (NVIDIA, Santa Clara, CA).

2.3. Pre-processing
A primary challenge in training a generative model to solve an image-to-image translation task such as this is constructing a set of training data consisting of well-matched pairs of images. This process becomes even more complicated in the abdomen when the variable presence of intestinal gas must be considered. Corresponding MRI and CT scans used in treatment planning may demonstrate notable mismatches in bowel filling and position that present a barrier to accurate dose calculations. While these mismatches could be handled during the treatment planning process in the clinical setting through intensive manual contouring to enable electron density overrides, we opt to avoid this entirely through the creation of an intensity projection prior. Here we adopt a novel approach to data augmentation for paired data DL applications in which incompatible
representations of intestinal gas in corresponding MRI and CT scans are made compatible through the propagation of air from MRI to CT images. The handling of intestinal gas proceeded in the following steps:

1. Corresponding MRI and CT scans were rigidly registered in the ViewRay treatment planning system (ViewRay Inc., Oakwood Village, OH) to achieve a gross alignment, primarily of bony anatomy.

2. Using automated thresholding and morphological reconstruction operations, regions of air in each scan were identified. For CT images, Otsu’s method (Otsu 1979) was used to compute a single threshold value with which the image could be quantized to produce a body mask that excludes most notably the couch. Within this mask, regions of air were thresholded with a histogram shape-based method by selecting intensities falling within an offset (defined as 7 bin widths here, nominally 140 HU) around the lower mode. Finally, binary erosion and dilation operations were performed to eliminate small, noisy regions and produce a smoother segmentation, respectively. For MR images, a similar approach was taken to produce a body mask. In identifying regions of air within this mask, extra precautions were required to avoid selecting low-signal regions not containing air (e.g. vertebral body). To this end, the erosion of a quantized image based on a five-level thresholding achieved with Otsu’s method was used as the basis for the morphological reconstruction (Vincent 1993) of segmented regions within the abdominal cavity as defined by the space surrounded by the body wall and excluding the region around the vertebral body.

3. These regions in the CT images were infilled with realistic texture via harmonic inpainting to produce a CT image with no air-containing regions (Schönlieb 2015, Parisotto and Schönlieb 2016).

4. Regions of air identified in the MR images in step (2) are then propagated to the corresponding CT images. As such, regions of air originally represented in the CT image but not the corresponding MR image maintain the infilled texture from step (3). Finally, the scans are deformably registered by way of a statistical dependence measure algorithm (Shi 2013).

2.4. Model and loss formulation

The task of sCT reconstruction viewed as a forward mapping from MRI to CT has been previously discussed (Olberg et al 2019). Briefly, the goal in establishing a generative model is to estimate a suitable operator that maps from MRI to CT, which is challenged by the many-to-one correspondence of pixel intensities between the two modalities. We approach the image-to-image translation task using a generative adversarial network framework consisting of two competing networks: (1) a generative model $G$ that produces CT samples residing in the same space as true CT data and (2) a discriminator $D$ that attempts to distinguish between samples generated by $G$ and true samples. During training, $G$ and $D$ undergo alternating minimization steps of their respective loss function, each of which depends on the generalized definition of sigmoid cross entropy loss (Abadi et al 2015):

$$L = \bar{x} - \bar{x} \ast \bar{x} + \log(1 + \exp(-\bar{x})),$$

(1)

where $\bar{x}$ is the true or predicted image logits computed by $D$ and $\bar{x}$ is a label corresponding to true ($\bar{1}$) or predicted ($\bar{0}$). Using this definition of sigmoid cross entropy loss, the generator loss function $g_{\text{loss}}$ is defined by

$$g_{\text{loss}} = L_{\text{adv}} + L_{\text{mae}},$$

(2)

where the adversarial loss $L_{\text{adv}}$ is the sigmoid cross entropy loss (equation (1)) with predicted image logits $\bar{x}$ assigned a true label ($\bar{z} = 1$) and the mean absolute error (MAE) loss $L_{\text{mae}}$ is the mean of the absolute difference between true images $I_{\text{true}}$ and predicted images $I_{\text{pred}}$:

$$L_{\text{mae}} = \text{mean}(|I_{\text{true}} - I_{\text{true}}|).$$

(3)

Minimizing such a loss formulation yields synthetic images that are computed as true images by the discriminator through the adversarial term $L_{\text{adv}}$, while also maintaining pixel-wise agreement between the generated and true images through the MAE term $L_{\text{mae}}$.

While true labels are assigned to predicted images in $g_{\text{loss}}$, the discriminator aims to correctly identify true and predicted images. As such, the discriminator loss function $d_{\text{loss}}$ depends only on the sigmoid cross entropy loss:

$$d_{\text{loss}} = L_{\text{pred}} + L_{\text{true}},$$

(4)

where $L_{\text{true}}$ and $L_{\text{pred}}$ are the sigmoid cross entropy loss (equation (1)) with true or predicted image logits $\bar{x}$ and corresponding labels $\bar{z} = 1$ or $\bar{z} = 0$, respectively. While $G$ strives to generate outputs computed as true images by $D$ through the adversarial loss term $L_{\text{adv}}$ with predicted image logits assigned a true label, the expected true and false labels are instead used for the true and predicted CT images in the formulation of $d_{\text{loss}}$.

Trainable parameters describing the various operations in each of the competing networks are optimized through alternating minimization steps of the respective loss functions. For $d_{\text{loss}}$, TensorFlow’s (Abadi et al 2015)
The gradient descent optimizer is used with an initial learning rate of $0.0002$. In the following minimization step, $E_{\text{loss}}$ is minimized utilizing the Adam gradient-based stochastic optimization algorithm (Kingma and Ba 2017) with an initial learning rate of $0.0002$, $\beta_1 = 0.7$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$. In both cases, learning rates decay every 10,000 steps subject to a staircase exponential function with a decay rate of 0.95.

2.5. Network architecture

2.5.1. Generator

The fully convolutional DenseNet (Jégou et al. 2017) employed here, illustrated in figure 1, consists of individual dense blocks arranged to form a stacked encoder-decoder U-net (Ronneberger et al. 2015) structure. The constituent dense blocks, which resemble residual blocks (He et al. 2015) in that intermediate feature maps are iteratively concatenated, are built using the following components: Batch Normalization (BN), ReLU activation, $3 \times 3$ same convolution, and dropout with probability $p = 0.2$. The growth rate $k$ of the layer ($k = 16$ in this case) dictates the number of feature maps computed by each layer. Feature maps computed by each of these intermediate layers are iteratively concatenated to form the output of the dense block itself, granting a degree of convergence-aiding supervision due to the short paths to all feature maps in an architecture that is ultimately quite parameter efficient (Jégou et al. 2017). Transition down (TD) operations in the encoder path, which serve to reduce the dimensionality of feature maps, consist of BN, followed by ReLU activation, $1 \times 1$ convolution, dropout with probability $p = 0.2$, and $2 \times 2$ max pooling with stride 2. To recover the input dimensions on the decoder path, transition up (TU) layers perform $3 \times 3$ transpose convolution with stride 2. Skip connections between corresponding layers of the encoder and decoder sides of the network transfer structural information that aids in the reconstruction of fine detail as the full input resolution is recovered along the decoder path.

2.5.2. Discriminator

The architecture of the discriminator is unchanged from the previous application to sCT generation in the breast (Olberg et al. 2019). $D$ is a straightforward encoder consisting of five convolutional layers that ultimately applies the sigmoid function to yield the probability of the evaluated image being a true CT image.

2.6. Evaluation

The proposed approach to sCT reconstruction is evaluated in two primary ways, each with a focus on the two classes of patients previously discussed. Pixel-wise image comparisons are made between true CT images and reconstructed sCT images for patients belonging to each class using the MAE and mean absolute percentage error (MAPE) measured in regions within the body contour not containing air:

$$\text{MAE} = \frac{\sum_{i=1}^{n} |CT_i - sCT_i|}{n}, \text{ and,}$$

$$\text{MAPE} = \frac{\sum_{i=1}^{n} \left| \frac{CT_i - sCT_i}{CT_i} \right|}{n}.$$
Completing 1500 epochs during training required 126 h in total. At deployment, inference requires 3.1 s.

### 3. Results

#### 3.1. Image comparison

These image comparisons are made for sCT outputs of both the model proposed here and a ‘blind’ model trained with the same patient data only without the pre-processing treatment of regions of air, which represents the conventional DL-based approach to the present problem. Considering the fact that the reference CT image for patients of Class 2 may be largely incompatible with the corresponding MRI image, we also evaluate the degree of overlap of regions of air in input MR images and reconstructed sCT outputs of each model using the Dice similarity index (DSC):

\[
\text{DSC} = \frac{2|X \cap Y|}{|X| + |Y|},
\]

where \(X\) and \(Y\) are the sets of pixels in air masks of an MR image and corresponding sCT image. Additionally, the structural similarity index (SSIM), which assesses similarity through three distinct luminance, contrast, and structure terms, (Olberg et al 2019) is calculated between the input MR image and reconstructed sCT output of each model in these air-containing regions.

A subsequent dosimetric evaluation compares dose distributions calculated at the first treatment fraction in simulation CT-based clinical plans to those recalculated using sCT-derived electron density information and the same optimization parameters used in the clinical plans for each of the 33 test patients. Full dose-volume histograms (DVHs) for the target and surrounding tissues of well-matched patients in Class 1 are used to first establish the baseline accuracy of the proposed approach to sCT reconstruction. The same comparison is made for patients in Class 2 demonstrating notable discrepancies in the presence of intestinal gas between corresponding MRI and CT scans to explore the effect of these discrepancies. For both patient classes, we examine differences in prescribed dose coverage of the PTV between the clinical CT-based plans and the proposed sCT-based plans. The 3D gamma index with a 3%/3 mm criterion is computed to evaluate agreement between the CT-based and sCT-based dose distributions for patients in each class (Low et al 1998). Additionally, mean DVH differences for each structure of interest are computed and evaluated for statistical significance in each patient class. For patients of Class 2, the full width at half maximum (FWHM) of the profile of the difference in calculated target coverage is used to evaluate the uncertainty in high-dose coverage of the target due to the involvement of intestinal gas. Dose calculations in each case were performed using the ViewRay treatment planning system integrated Monte Carlo algorithm in the presence of a magnetic field with a dose grid resolution of 0.3 cm and calculation uncertainty of 1%.
0.56 ± 0.22 for the blind model to 0.80 ± 0.21 for the proposed model. In these air-containing regions, the SSIM computed between the input MR image and reconstructed sCT image improved from 0.14 ± 0.06 to 0.34 ± 0.07 with the adoption of the proposed model over the blind model.

Figure 2. Image comparisons for representative slices of well-matched patients of Class 1. Input MR images (a), output sCT images for the blind model (b) and proposed model (c), true CT images (d), and absolute difference maps (e)–(f) between the true CT images and generated sCT images for the blind and proposed model, respectively. Values in the absolute difference maps are in units of HU. The final row illustrates the rare case of a relatively well-matched slice with a notable presence of intestinal gas.

Figure 3. Image comparisons for representative slices of patients in Class 2 characterized by notable differences in the presence of intestinal gas between corresponding MRI and CT scans. Input MR images (a), output sCT images for the blind model (b) and proposed model (c), true CT images (d), and absolute difference maps (e)–(f) between the true CT images and generated sCT images for the blind and proposed model, respectively. Values in the absolute difference maps are in units of HU.
Similarly, the dosimetric evaluation of the proposed approach to sCT reconstruction focused on the two distinct classes of patients. In both cases, optimization parameters selected in the CT-based clinical plans were used to recalculate dose distributions based on electron density information derived from the generated sCT images. Used as a baseline point of reference to establish the dosimetric accuracy of the proposed reconstructions, the thirteen well-matched patients of Class 1 demonstrate differences in the prescribed dose coverage of the PTV ($V_{100}$) of $1.3 \pm 2.1\%$ between CT-based clinical plans and the sCT-based plans with a gamma pass rate of $98.3 \pm 1.3\%$ using 3%/3 mm criterion. The representative DVHs for patients belonging to Class 1 shown in figure 4 demonstrate close agreement in calculated target coverage and doses to surrounding tissues between the CT-based clinical plans and the sCT-based recalculations.

For the twenty poorly-matched patients of Class 2, notable discrepancies in the representation of intestinal gas between corresponding MRI and CT scans result in sizeable and variable differences in PTV $V_{100}$ coverage: $13.3 \pm 11.0\%$ on average. Due to these differences, the gamma pass rate is reduced to $93.9 \pm 9.8\%$ using the same 3%/3 mm criterion. These differences in target coverage, along with small discrepancies in the dose to closely involved tissues like the duodenum, are observable in the representative DVHs included in figure 5.
plotted in figure 5 is the difference in target coverage at each point, which yields an approximately Gaussian profile. The FWHM of this profile for all patients in Class 2 covers an average range of 51.4–58.2 Gy. Mean DVH differences between the CT-based clinical plans and the sCT-based plans for all structures of interest are plotted in figure 6 for patients of each class. For patients in Class 1, differences between the CT-based and sCT-based plans are computed to be statistically insignificant (distributed with a median of zero) using a two-sided Wilcoxon signed rank test (Wilcoxon 1945) for each of the duodenum ($p = 0.34$), large bowel ($p = 0.62$), liver ($p = 0.52$), small bowel ($p = 0.38$), spinal cord ($p = 0.91$), stomach ($p = 0.91$), and PTV ($p = 0.20$). For patients in Class 2, differences were instead shown to be statistically significant (not distributed with a median of zero) using a two-sided Wilcoxon signed rank test for each of the duodenum ($p = 0.002$), large bowel ($p = 0.007$), liver ($p < 0.001$), small bowel ($p < 0.001$), spinal cord ($p = 0.01$), stomach ($p < 0.001$), and PTV ($p < 0.001$).
4. Discussion

The potential value of generating synthetic CT data for MRI-only ART in the abdomen is multifaceted. Although therapeutic gains may be achieved, adopting an online adaptive workflow introduces additional time burdens to the process of treatment delivery including re-contouring, re-planning, and quality assurance—all of which must occur while the patient remains on-table. Re-contouring, which must be undertaken to accommodate changes in normal tissue volumes and—ideally—the variable presence of intestinal gas that is our current focus, represents a significant portion of the total on-table time per fraction: up to 24 min in the worst case (Henke et al 2018). By utilizing the proposed approach to sCT reconstruction explored here in which the focus was placed on producing a clinically unavailable data set of well-matched representations of intestinal gas, one is able to rapidly produce (0.26 s/slice) sCT data in the clinical setting that accurately reflects both the presence of intestinal gas shown in a patient’s daily MRI scan and HU values present in a true CT scan. In this way, one of the primary concerns prompting re-contouring in the adaptive setting is potentially eliminated. When paired with an auto-contouring strategy designed for MRI-guided ART (Fu et al 2018), the time burden associated with re-contouring may become negligible.

Another primary motivation in adopting an adaptive workflow is to achieve dose escalation under shifting anatomic conditions (Henke et al 2018). Plan adaptation is often performed to increase OAR sparing while also increasing target coverage (Henke et al 2018). The observed underdosing of the target for patients of Class 2 characterized by mismatched representations of intestinal gas is especially relevant in these scenarios when dose escalation is a specific aim of pursuing plan adaptation. The higher calculated target coverage in clinical plans in which mismatches between the planning CT and setup MRI are not accounted for compared to the sCT-based plans and uncertainty in high-dose regions (figure 5) represent a barrier to any escalation that is pursued in these scenarios. In the present study, we have explored this effect at the first treatment fraction for 33 test patients, demonstrating that even at the first fraction, a non-negligible portion of the patient population may experience uncertainties in simulation CT-based dose calculations due to the involvement of intestinal gas. This concern becomes even more important when considering the accumulation of dose over the course of a treatment in which each fraction is adapted and these differences accrue. However, it is important to note that the DVHs computed for a given treatment fraction and presented in the ViewRay treatment planning system are full rather than fractional DVHs. Considering this, the DVHs examined in this study convey the overall effect in the case that the magnitude of the discrepancies observed at the first fraction are carried forward through each subsequent fraction. Considering the hard-to-characterize nature of gas motion, we do not examine here whether there exists some sort of interplay effect throughout the course of treatment.

We acknowledge a number of additional limitations to the present study. First and foremost, the image-to-image translation approach employed here struggles fundamentally in a situation in which the anatomy represented in corresponding training image pairs differs. Although we took care in the present study to propagate regions of air, geometric differences in soft tissues surrounding the target are not always handled sufficiently by multi-modal deformable image registration. These general effects are observed in image...
comparisons, where outputs of the blind model appear to be uniformly lower in intensity compared to the true CT images. We attribute this to the increased ambiguity in pixel correspondences in the training image pairs. In effect, more pixel intensities are pushed towards the lower intensity of air. In the proposed case with the pre-processed training data set, the limiting ambiguity becomes the difference in soft tissue representation in the corresponding MR and CT images. The model is reasonably robust to the variations present in the training data set, but struggles to faithfully reconstruct the most dynamic tissues and structures. Style-transfer methods that do not rely on matched pairs of training data as exemplified by CycleGAN (Zhu et al 2018) may be of particular use in this application to overcome the limitations of multi-modal deformable image registration. The implementation of an unpaired approach may still benefit from semi-supervised data produced in the manner described here to overcome the limitations of the unpaired approach in cases with inherent ambiguity, as is the case for MRI signal intensity. A second limitation of the trained model is reconstruction inaccuracies at the superior and inferior extremes of a patient’s image stack that stem from the make-up of the training data set. While every patient data set was roughly centered on the target and surrounding tissues, slices containing views of the lungs and diaphragm or inferior portions of the abdomen were not as equally represented. These reconstruction inaccuracies may be ameliorated by adopting a more robust 3D network architecture over the relatively lightweight 2D architecture at the expense of heavily increased memory usage, which may not be a feasible trade-off in all settings. An additional concern regarding MR image quality in the abdominal sCT reconstruction task is the issue of susceptibility artifacts in the GI tract. The effects of these artifacts are lessened due to the lower field strength of the MRgRT platform utilized here, (Ginn et al 2017) but certain circumstances involving the ingestion of fortified foods prior to treatment—although not encountered in the present study—have been reported and are thus an important consideration in this application (Green et al 2018).

An inherent challenge in this space is the issue of image evaluation when the underlying premise is that the ‘ground truth’ simulation CT data is incompatible with the setup MRI data used due to the involvement of intestinal gas and the motion of GI structures. In computing both the MAE and MAPE, direct pixel-wise comparisons are made between the sCT images and the corresponding CT images. In a similar way, the multi-modality image similarity comparison made between MR images and sCT reconstructions by way of the SSIM is challenged by modality-specific differences in contrast and luminance. As such, the image comparisons made here are imperfect comparisons and we instead rely more heavily on the improvement in the representation of air in our sCT reconstructions as measured by the DSC considering that is the primary focus of this work. Nonetheless, a comparison to other methods is warranted. Ahunbay et al (2019) achieved an MAE of approximately 25 HU in the abdomen using a method entirely reliant of multiple deformations of true CT images. Closer to the realm of true image synthesis, multiple atlas-based techniques have reported values ranging from 40 to 200 HU for sites including the pelvis, cranium, and general torso (Edmund and Nyholm 2017, Johnstone et al 2018, Guerreiro et al 2019). Finally, results of DL-based approaches at various sites quantified using various metrics have been collected and reported by Spadea et al (2021) for general comparison with the results achieved here. For abdominal cases demonstrating little involvement of intestinal gas, (Cusumano et al 2020) report an MAE of 78.71 ± 18.46 within the body contour. In a similar cohort, (Qian et al 2020) report MAE values ranging from approximately 42–79 HU for two DL-based methods. In the MAE values we report, we do not distinguish between regions of soft tissue and bone, but we do compare competitively to existing methods.

Finally, the separation of test patients into two separate classes performed here relied on the qualitative assessment of the relative involvement of intestinal gas and the degree to which representations of intestinal gas matched between corresponding MRI and CT scans. In some cases, the characterization of the patient was clear—there were easily observable discrepancies or gas was entirely uninvolved—but the categorization was more challenging in other cases. As such, the furthering of this and related work would benefit from some quantitative approach to the characterization of the involvement and similarity of representations of intestinal gas that would in turn enable the exploration of trends in distinct patient groups.

5. Conclusions

The approach to sCT reconstruction in the abdomen evaluated here highlights the challenges posed by the presence of intestinal gas throughout the MRI-guided ART workflow. Eliminating the burden of handling intestinal gas from the clinical setting through the creation of a clinically unavailable training data set for training a paired data generative model offers the potential to streamline a time-intensive portion of the adaptive treatment workflow. These time savings are gained while also enabling accurate dose calculations in adaptive treatments despite the variable presence of intestinal gas at each stage of treatment planning and delivery during MRI-only ART in the abdomen.
Acknowledgments

JSDK acknowledges this work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. 2020R1A4A101661911).

Conflicts of interest

The authors have no conflicts to disclose.

ORCID iDs

Jaehee Chun  https://orcid.org/0000-0002-9695-6079
Taeho Kim  https://orcid.org/0000-0001-9809-9587

References

Abadi M et al 2015 TensorFlow: Large-scale machine learning on heterogeneous systems. 2015. Software available from https://www.tensorflow.org/about/bib
Ahunbay E E, Thapa R, Chen X, Paulson E and Li X A 2019 A technique to rapidly generate synthetic computed tomography for magnetic resonance imaging-guided online adaptive replanning: an exploratory study Int. J. Radiat. Oncol. 103 1261–70
Arabi H et al 2018 Comparative study of algorithms for synthetic CT generation from MRI: consequences for MRI-guided radiation planning in the pelvic region Med. Phys. 45 5218–33
Bredfeldt J S, Liu L, Feng M, Cao Y and Balter J M 2017 Synthetic CT for MRI-based liver stereotactic body radiotherapy treatment planning Phys. Med. Biol. 62 2922–34
Chen S, Qin A, Zhou D and Yan D 2018 Technical note: U-net-generated synthetic CT images for magnetic resonance imaging-only prostate intensity-modulated radiation therapy treatment planning Med. Phys. 45 5659–65
Corradi N et al 2019 MR-guidance in clinical reality: current treatment challenges and future perspectives Radiat. Oncol. 14 92
Cusumano D et al 2020 A deep learning approach to generate synthetic CT in low field MR-guided adaptive radiotherapy for abdominal and pelvic cases Radiother. Oncol. 153 205–12
Ding S and Nyholm T 2017 A review of substitute CT generation for MRI-only radiation therapy Radiat. Oncol. 12 28
Fallone B G 2014 The rotating biplanar linac-magnetic resonance imaging system Semin. Radiat. Oncol. 24 200–2
Feng M et al 2009 Characterization of pancreatic tumor motion using cine MRI: surrogates for tumor position should be used with caution Int. J. Radiat. Oncol. Biol. Phys. 74 884–91
Fischer-Valkuck B W et al 2017 Two and a half-year clinical experience with the world’s first magnetic resonance image guided radiation therapy system Adv. Radiat. Oncol. 2 485–93
Fu J et al 2020 Generation of abdominal synthetic CTs from 0.35 T MR images using generative adversarial networks for MR-only liver radiotherapy BioMed. Phys. Eng. Express 6 015033
Fu Y et al 2018 A novel MRI segmentation method using CNN-based correction network for MRI-guided adaptive radiotherapy Med. Phys. 45 5129–37
Ginn J et al 2017 Characterization of spatial distortion in a 0.35 T MRI-guided radiotherapy system Phys. Med. Biol. 62 4525–40
Green O, Henke L, Parikh P, Roach M, Michalski J and Gach H 2018 Practical implications of ferromagnetic artifacts in low-field MRI-guided radiotherapy Curies 10 1–6 e2359
Guerreiro F et al 2019 Feasibility of MRI-only photon and proton dose calculations for pediatric patients with abdominal tumors Phys. Med. Biol. 64 1–13
Han X 2017 MR-based synthetic CT generation using a deep convolutional neural network method Med. Phys. 44 1408–19
He K, Zhang X, Ren S and Sun J 2015 Deep residual learning for image recognition arXiv:1512.03385
Henke L et al 2018 Phase I trial of stereotactic MR-guided online adaptive radiation therapy (SMART) for the treatment of oligometastatic or unresectable primary malignancies of the abdomen Radiother. Oncol. 126 519–26
Henke L et al 2018 Magnetic resonance image-guided radiotherapy (MRigRT): a 4.5 year clinical experience Clin. Oncol. 30 720–7
Hsu S H, Peng Q and Tomé W A 2019 On the generation of synthetic CT for a MRI-only radiation therapy workflow for the abdomen J. Phys.: Conf. Ser. 1154 1–4
Jégou S, Drozdziel M, Vazquez D, Romero A and Bengio Y 2017 The one hundred layers tiramisu: fully convolutional DenseNets for semantic segmentation arXiv:1611.09326 1–9
Johnstone E et al 2018 Systematic review of synthetic computed tomography generation methodologies for use in magnetic resonance imaging-only radiation therapy Int. J. Radiother. Oncol. 100 199–217
Keall P J, Barton M and Crozier S 2014 The Australian magnetic resonance imaging-linac program Semin. Radiat. Oncol. 24 203–6
Kimura D and Bai J 2017 Adam: a method for stochastic optimization arXiv:1412.6980 1–15
Koivula L, Wei L and Korhonen J 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
Koivula L et al 2017 Intensity-based dual model method for generation of synthetic CT images from standard T2-weighted MR images—generalized technique for four different MR scanners Radiat. Oncol. 125 411–9
Korhonen J, Kapanen M, Kyyriläinen 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 011704
Kumagai M et al 2009 Impact of intrafractional bowel gas movement on carbon ion beam dose distribution in pancreatic radiotherapy Int. J. Radiat. Oncol. Biol. Phys. 73 1276–81
Low D A, Harms W B, Matic S and Purdy J A 1998 A technique for the quantitative evaluation of dose distributions Med. Phys. 25 656–61
Maspero M et al 2018 Dose evaluation of fast synthetic -CT generation using a generative adversarial network for general pelvis MR-only radiotherapy Phys. Med. Biol. 63 1–11


Mostafaei F et al 2018 Variations of MRI-assessed peristaltic motions during radiation therapy PLoS One. 13 e0205917
Mucic S and Dempsey J F 2014 The ViewRay system: magnetic resonance-guided and controlled radiotherapy Semin. Radiat. Oncol. 24 196–9
Nakamoto Y, Chin B B, Cohade C, Osman M, Tatsumi M and Wahl R L 2004 PET/CT: artifacts caused by bowel motion Nucl. Med. Commun. 25 221–5
Noel C E et al 2015 Comparison of onboard low-field magnetic resonance imaging versus onboard computed tomography for anatomy visualization in radiotherapy Acta. Oncol. 54 1474–82
Olberg S et al 2019 Synthetic CT reconstruction using a deep spatial pyramid convolutional framework for MR-only breast radiotherapy Med. Phys. 46 4135–47
Otsu N 1979 A threshold selection method from gray-level histograms IEEE Trans. Syst. Man Cybern. 9 62–6
Parisotto S and Schönlieb C (2016) MATLAB/Python codes for the Image Inpainting Problem zenodo.org/record/4315174
Peng Y et al 2020 Magnetic resonance-based synthetic computed tomography images generated using generative adversarial networks for nasopharyngeal carcinoma radiotherapy treatment planning Radiat. Oncol. 150 217–24
Pollard J M, Wen Z, Sadagopan R, Wang J and Ibbott G S 2017 The future of image-guided radiotherapy will be MR guided Br. J. Radiol. 90 20160667
Qian P et al 2020 Estimating CT from MR abdominal images using novel generative adversarial networks J. Grid Comput. 18 211–26
Raaymakers B W et al 2017 First patients treated with a 1.5 T MRI-Linac: clinical proof of concept of a high-precision, high-field MRI guided radiotherapy treatment Phys. Med. Biol. 62 L41–50
Ronneberger O, Fischer P and Brox T 2015 U-net: convolutional networks for biomedical image segmentation arXiv:1505.04597
Schmidt M A and Payne G S 2015 Radiotherapy planning using MRI Phys. Med. Biol. 60 R323–61
Schönlieb C B 2015 Partial Differential Equation Methods for Image Inpainting (Cambridge, UK: Cambridge University Press)
Shi J 2013 Statistical dependence measure based multi-modal image registration and registration assisted non-parametric image segmentation PhD Thesis University of Florida
Spadea M F, Maspero M, Zaffino P and Seco J 2021 Deep learning-based synthetic-CT generation in radiotherapy and PET: a review arXiv:2102.02734
Vincent L 1993 Morphological grayscale reconstruction in image analysis: applications and efficient algorithms IEEE Trans. Image Process. 2 176–201
Wen N et al 2018 Magnetic resonance guided linear accelerator for stereotactic radiosurgery treatment Int. J. Radiat. Oncol. 102 e479
Wilcoxon F 1945 Individual comparisons by ranking methods Biometrics Bull. 1 80–3
Zhu J Y, Park T, Isola P and Efros A A 2018 Unpaired image-to-image translation using cycle-consistent adversarial networks arXiv:1703.10593