Article Type: Research Article

Generating Synthetic CTs from Magnetic Resonance Images using Generative Adversarial Networks

Hajar Emami[1], Ming Dong[1], Siamak P. Nejad-Davarani[2], Carri Glide-Hurst[2]

1. Department of Computer Science, Wayne State University, Detroit, Michigan 48202, USA.

2. Department of Radiation Oncology, Henry Ford Health System, Detroit, Michigan 48202, USA.

Corresponding authors:

Carri Glide-Hurst, churst2@hfhs.org

Ming Dong, ak3389@wayne.edu

Running title: SynCT Generation via GAN

ABSTRACT

Purpose: While MR-only treatment planning using synthetic CTs (synCTs) offers potential for streamlining clinical workflow, a need exists for an efficient and automated synCT generation in the brain to facilitate near real-time MR-only planning. This work describes a novel method for

generating brain synCTs based on generative adversarial networks (GANs), a deep learning model that trains two competing networks simultaneously, and compares it to a deep convolutional neural network (CNN).

Methods: Post-Gadolinium T1-Weighted and CT-SIM images from fifteen brain cancer patients were retrospectively analyzed. The GAN model was developed to generate synCTs using T1-weighted MRI images as the input using a residual network (ResNet) as the generator. The discriminator is a CNN with five convolutional layers that classified the input image as real or synthetic. Five-fold cross validation was performed to validate our model. GAN performance was compared to CNN based on mean absolute error (MAE), structural similarity index (SSIM), and peak signal-to-noise ratio (PSNR) metrics between the synCT and CT images.

Results: GAN testing took ~11 hours with a new case testing time of 5.7±0.6 seconds. For GAN, MAEs between synCT and CT-SIM were 89.3±10.3 HU and 41.9±8.6 HU across the entire FOV and tissues, respectively. However, MAE in the bone and air was, on average, ~240-255 HU. By comparison, the CNN model had an average full FOV MAE of 102.4±11.1 HU. For GAN, the mean PSNR was 26.6±1.2 and SSIM was 0.83±0.03. GAN synCTs preserved details better than CNN, and regions of abnormal anatomy were well represented on GAN synCTs.

Conclusions:

We developed and validated a GAN model using a single T1-weighted MR image as the input that generates robust, high quality synCTs in seconds. Our method offers strong potential for supporting near real-time MR-only treatment planning in the brain.

Key Words: Generative adversarial network, synthetic CTs, radiation therapy

Introduction

Because magnetic resonance imaging (MRI) has excellent soft tissue contrast that substantially improves target and organ at risk (OAR) segmentation accuracy and reliability [1-4], it is currently used as an adjunct to CT-based radiation therapy planning (RTP). However, rigid MRI-CT co-registration has been reported to introduce geometrical uncertainties of 0.5 to 2 mm for the brain [5-7]. To circumvent these uncertainties and streamline clinical efficiency (eliminating CT scans reduces dose, patient time, and costs), recent attention has been given to MR-only treatment planning in the brain using synthetic CTs (synCTs) generated from MRI data [8-14]. Many techniques have been proposed for generating synCTs. Atlas-based methods typically consist of performing a deformable registration between a previously developed single or multiple atlases and a test MR image, and estimating an attenuation map via image warping [15-17]. A drawback of atlas-based methods is that accuracy depends upon the deformable registration performance which is challenging to validate in multi-modality workflows. Ultra-short echo time (UTE) sequences are often incorporated to help improve visualization and separation of tissues with short T2 such as air and bone[18]. Image processing pipelines have been developed incorporating magnitude[9, 13, 17, 19] and phase[12] UTE data in combination with other MRI sequences. However, UTE sequences are not widely available, and thus, efforts to generate synCT from clinical sequences, or preferably a single sequence, commonly acquired for RTP are advantageous.

Recently, Han developed a learning-based approach in which Convolutional Neural Networks (CNNs) are employed to perform image-to-image mapping between T1-weighted MR and CT datasets[20]. Once trained, the model was found to outperform both atlas-based and regression-based approaches for generating synCTs. Experimental results have shown that the number of layers (i.e., CNN depth) has a significant impact on the CNN performance in image

recognition tasks, suggesting that deeper networks have better overall performance[21]. However, creating deeper CNNs by simply adding more layers to the architecture may lead to what is termed a degradation problem, or when the network accuracy gets saturated by increasing the depth of the network and then has been shown to rapidly degrade[22]. To address this degradation, a deep residual network (i.e., ResNet) has been introduced that adds shortcut connections, thereby skipping one or more layers without adding extra parameters or computational complexity[22].

Generative adversarial networks (GANs) [23] are another type of deep learning model that trains two competing networks simultaneously: a generator network to synthesize data and a discriminator network to distinguish between the synthesized and real data. GAN has been shown to generate realistic results such as predicting objects from sketches and color from gray-scale images [24] however medical imaging applications are currently limited. Nie *et al*. recently presented a 3D fully convolutional network (FCN) as the GAN generator with an additional discriminator network to improve the realistic nature of synthetic CTs derived from MRI data in the brain and pelvis. The method was found to outperform an atlas-based approach and an FCN model[25].

In this work, we introduce a novel approach for generating synCTs from T1-weighted post-Gadolinium MRI datasets by applying a GAN model with a ResNet architecture as the generator. To quantify potential advantages of implementing GAN, we then compared our results to a trained deep convolutional neural network (CNN). Our experimental results show that the proposed GAN model can efficiently and accurately generate synCT images from the MRI input while outperforming CNN, thus offering strong potential for supporting MR-only radiation therapy workflows.

Materials and Method

GAN Network Architecture

Figure 1 outlines the proposed GAN model, consisting of ResNet and regular CNN architectures as the generator and discriminator, respectively. The ResNet generator architecture has shortcut connections (shown as solid arrows in Figure 1) for each ResNet block that skip one or more layers. These shortcut connections are identity maps used for adding the input of each block to the output. This methodology eases the training of the deep network without adding extra parameters or computational complexity [22].

The first three convolutional layers of the ResNet architecture have 64, 128 and 256 number of filters with kernel sizes of 7×7, 3×3 and 3×3, respectively. Filters are applied with a stride of 2 (filters convolve around the input by shifting 2 units at a time). Each convolutional layer uses its filters to perform 2D convolutions on its input. The input of each layer is the output of its previous layer, except for the first layer in which the input is the MR images.

The operation of the convolutional layers can be expressed as:

$$h_k = W_k * X + b_k, k \in [0, k-1] \quad (1)$$

where W_k and b_k are the weights and bias of the $k-th$ filter, respectively, and '$*$' is the convolution operation. The output of Eq. (1) is the $k-th$ channel of the output feature map of this convolutional layer. If the input feature map has size (H_{in}, W_{in}), the spatial size of the output feature map can be computed as a function of the input volume size (H_{in}, W_{in}), the receptive field size of the convolutional layer neurons F, the stride with which they are applied S, and the amount of zero padding P on the border:

$$H_{out} = floor\left(\left(H_{in} - F + 2*P\right)/S + 1\right)$$
$$W_{out} = floor\left(\left(W_{in} - F + 2*P\right)/S + 1\right) \quad (2)$$

The convolution outputs are followed by a batch normalization layer and a nonlinear activation function: Rectified Linear Unit (ReLU). Batch normalization allows using much higher learning rates and accelerates the network training[26]. The batch normalization transform over a mini batch can be expressed as:

$$y_i = \frac{x_i - \mu_B}{\sqrt{\delta_B^2 + \varepsilon}} * \gamma + \beta \quad (3)$$

The mean and standard deviation are calculated per dimension over the mini batches, γ and β are learnable parameters, and ε is a value added to the denominator for numerical stability, normally set to 10^{-5}. The scaled and shifted values y are passed to other network layers. ReLU is a nonlinear activation function:

$$\mathrm{Re}LU(a) = \begin{cases} a, & if \quad a > 0 \\ \delta * a & otherwise \end{cases} \quad (4)$$

where a is an element in the feature map, and δ is a slope value to leak negative elements.

After the first three convolutional layers, we have nine residual blocks. In each residual block, there are two convolutional layers with 3×3 kernel and 256 filters followed by batch normalization layers and the ReLU activation function. Each residual block has a shortcut connection to add the input of that block to the output of the last convolutional layer in the block.

ResNet has a general form of residual blocks with fully connected layers. In our GAN model, the generator does not require a fully connected layer in its architecture. Since our goal is to generate synthetic CT images of input size, we also added two transposed convolutional layers

after the residual blocks. Transposed convolutional layers are used when a transformation in the opposite direction of a normal convolution is needed. For instance, such a transformation is usually needed in the decoding layer of a convolutional autoencoder or to project feature maps to a higher dimensional space. After nine residual blocks, we have two transposed convolutional layers with kernel size 3×3 that have 128 and 64 filters. Transposed convolutional layers are used when a transformation in the opposite direction of a normal convolution is needed. For instance, such a transformation is usually needed in the decoding layer of a convolutional autoencoder or to project feature maps to a higher dimensional space. If the input feature map of these layers has size (H_{in}, W_{in}), the spatial size of the output feature map can be computed as a function of the input volume size (H_{in}, W_{in}), the receptive field size of the convolutional layer neurons F, the stride with which they are applied S, and the amount of zero padding P on the border:

$$\begin{aligned} H_{out} &= (H_{in} - 1) * S - 2 * p + F \\ W_{out} &= (W_{in} - 1) * S - 2 * p + F \end{aligned} \tag{5}$$

These transposed convolutional layers are followed by batch normalization layers and the ReLU activation function. At the end, there is one convolutional layer with kernel size 7×7 followed by the activation function.

In addition, we use dropout in layers of the generator as an effective technique for regularization and preventing the co-adaptation of neurons in the network. Dropout randomly zeroes some of the elements of the input with probability p using samples from a bernoulli distribution. The dropout layer is defined as below:

$$\upsilon_{drop} = \begin{cases} \dfrac{\upsilon}{1-p}, & if \quad u > p \\ 0, & otherwise \end{cases} \tag{6}$$

where υ is the input, $u \sim U(0,1)$ is generated independently for each input at each iteration, p is the dropout ratio set to 0.5 (any given unit is dropped at the probability of 0.5), υ_{drop} is the output after dropout.

The discriminator network of our GAN model is a CNN with five convolutional layers followed by batch normalization and ReLU that classifies an input image as real or synthetic. The discriminator can help the generator network to produce the more realistic synCT output from the MRI input. All convolutional layers in the discriminator have a filter size 4×4 with stride 2.

Objective

This work employs a conditional GAN model for generating synCTs[24]. A conditional GAN model learns a mapping from the input image *x* (an MR image) to the output image *y* (a real CT image). In our study, the model learns mapping from MR images to CT images. We replace the negative log likelihood objective by a least squares loss which has been shown to be more stable during training and generates higher quality results [27, 28].

The discriminator (*D*) loss ℓ_D and the generator (*G*) loss ℓ_G of our GAN model are defined as:

$$\ell_D = \frac{1}{2}\mathrm{E}_{y\sim p_{data}(y)}\left[(D(y)-1)^2\right] + \frac{1}{2}\mathrm{E}_{x\sim p_{data}(x)}\left[D(G(x))^2\right] \quad (7)$$

and

$$\ell_G = \frac{1}{2}\mathrm{E}_{x\sim p_{data}(x)}\left[(D(G(x))-1)^2\right] + \lambda \ell_{L1}(G) \quad (8)$$

where *pdata* is the data probability distribution, $G(x,z)$ is the synCT image and D gives the probability of the input being a real CT image. Note that the L1 distance is used as a part of the reconstruction error to generate more realistic output images:

$$\ell_{L1}(G) = \mathrm{E}_{x,y\sim P_{data}(x,y)}\left[\|y - G(x)\|_1\right] \quad (9)$$

The optimal solution of our objective function is achieved under two conditions: 1) the discriminator is not able to distinguish between synCTs and real CTs, and 2) the difference between synCTs and real CTs is minimized. To evaluate the impact of loss function selection, we also compared results from implementing least squares loss and negative log likelihood.

Optimization

The detailed procedure of network optimization is presented in Algorithm 1. The generator is not trained until the discriminator has been trained for k steps (k=1 in our study).

Algorithm 1 GAN training procedure: One gradient descent step is applied on the discriminator and on the generator, alternatively. $\alpha = 0.001, \beta_1 = 0.9, \beta_2 = 0.999,$ and $\varepsilon = 10^{-8}$. β_1^t, and β_2^t are β_1 and β_2 to the power *t*.

Randomly initialize θ_d, θ_g

for number of training iterations **do**

- Sample minibatch of *m* examples $\{x^{(k)}\}_{k=1}^{m} \sim p_{data}(x)$, $\{y^{(k)}\}_{k=1}^{m} \sim p_{data}(y)$.
- Update θ_d at iteration *t*:
  - (1) $g_t \leftarrow \nabla_{\theta_d}\left(\frac{1}{m}\sum_{i=1}^{m}\left[\left(D(y^i)-1\right)^2 + D\left(G\left(x^i\right)\right)^2\right]\right)$
  - (2) $m_t \leftarrow \beta_1 . m_{t-1} + (1-\beta_1).g_t$ (Update biased first moment estimate)
  - (3) $v_t \leftarrow \beta_2 . v_{t-1} + (1-\beta_2).g_t^2$ (Update biased second raw moment estimate)
  - (4) $\hat{m}_t \leftarrow m_t / (1-\beta_1^t)$ (Compute bias-corrected first moment estimate)
  - (5) $\hat{v}_t \leftarrow v_t / (1-\beta_2^t)$ (Compute bias-corrected second raw moment estimate)
  - (6) $\theta_{d,t} \leftarrow \theta_{d,t-1} - \alpha . \hat{m}_t / \left(\sqrt{\hat{v}_t} + \varepsilon\right)$ (Update parameters)
- Update θ_g:

$$g_t \leftarrow \nabla_{\theta_g}\left(\frac{1}{m}\sum_{i=1}^{m}\left(D\left(G\left(x^i\right)\right)-1\right)^2+\lambda\left\|y^i-G\left(x^i\right)\right\|_1\right)$$

repeat above steps (2) to (6)

end for

Training Dataset and Preprocessing

Images from fifteen brain cancer patients were retrospectively analyzed as part of an IRB approved study. MR-SIM was performed on a 1.0T Panorama High Field Open (Philips Medical Systems, Best, Netherlands) using an 8-channel head coil. No immobilization devices were used during MR-SIM due to incompatibility of the thermoplastic mask in the head coil. Post-Gadolinium T1-weighted images were acquired for each patient with a voxel size of $0.90\times0.90\times1.25$ mm^3. Brain CT-SIM was performed on a Brilliance Big Bore (Philips Health Care, Cleveland, OH) scanner with the following parameters: 500 mAs, 0.88×0.88 mm^2 in-plane spatial resolution, and 1-mm slice thickness.

All MR images were then aligned to their corresponding CTs using Statistical Parametric Mapping (SPM) 12 (Functional Imaging Laboratory, Wellcome Trust Centre for Neuroimaging, UCL), a well-known registration software commonly used in neurology applications[29]. In-plane resolution and slice thickness were interpolated to match to the CT frame of reference. Normalized mutual information [30] was used as the objective function with trilinear interpolation method used for re-slicing. CT-SIM data were used as ground truth comparisons by excluding voxels outside of the external contour to eliminate the impact of immobilization devices and couch structures. In the MRI data, a binary head mask was derived from each MR slice using thresholding and morphological operators to separate the head region from the external devices. To validate our model's performance, a five-fold cross validation technique was used for training and testing steps in a similar manner to recent work published by Han[20] .

That is, 15 cases are randomly partitioned into five groups, and for each experiment, four groups (including twelve cases) are selected for training the model and one group (including three cases) is selected to test the trained model.

Comparison of CNN and GAN

Han developed a learning-based approach using CNN for generating synCTs, in which they adopted and used the U-net model [20]. U-net is a well-established FCN architecture with two symmetrical parts: an encoding part and a decoding part, originally proposed in [31]. Similar to a regular CNN, there are convolutional and pooling layers in the encoding part of the U-net architecture to extract the context of the input image. The decoding part includes convolutional and unpooling layers to reconstruct the image prediction from low to high resolution. They showed that CNN outperformed atlas-based and regression-based approaches in generating synCTs. In general, GAN has been shown to outperform CNN for image generation tasks[24, 32, 33]. CNN tries to minimize the Euclidean distance between the generated images and the ground truth, leading to blurry results because Euclidean distance is minimized when all possible outputs are averaged. On the other hand, GAN has a loss function that tries to distinguish between real or synthetic images and trains a generator and a discriminator to minimize that loss. Since the discriminator in GAN model can easily classify blurry images as the synthetic ones, blurry images produced by the generator will result in a higher loss[24]. Reference [33] highlighted that GAN generates more realistic images compared to CNN when evaluated using visual inspection. In the proposed model, the generator in the GAN receives the feedback of the discriminator through a back-propagation step that is expected to result in more accurate synCT generation. To evaluate this impact, following techniques developed by Han [20], we implemented and trained U-net with 27 convolutional layers with no discriminator

block to compare against the GAN approach for the metrics listed below. To further quantify the impact of using adversarial learning, comparisons were made between our GAN model, where the generator receives feedback from the discriminator through a back-propagation step, and a comparison to a ResNet model with no discriminator block.

Quantitative Measurements and Evaluation

Three metrics were used to evaluate the prediction accuracy of the model between the synCT and CT-SIM data: mean absolute error (MAE), structural similarity index (SSIM), and peak signal-to-noise ratio (PSNR) as defined below.

$$MAE = \frac{\sum_{i=1}^{H} |realCT(i) - synCT(i)|}{H} \tag{9}$$

$$SSIM = \frac{(2\mu_x\mu_y + C_1)(2\delta_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\delta_x^2 + \delta_y^2 + C_2)} \tag{10}$$

$$PSNR = 10\log_{10}(\frac{Q^2}{MSE}) \tag{11}$$

In Eq. (9), H is the number of head voxels. For the MAE metric, the lower value means better prediction results, i.e., more realistic synCTs. In Eq. (10), x is the real CT image, y is the corresponding synthetic CT image, μ_x denotes the mean value of image x, μ_y denotes the mean value of image y, δ_x^2 is the variance of image x, δ_y^2 is the variance of image y, and the parameters $C_1 = (k_1 Q)^2$ and $C_2 = (k_2 Q)^2$ are two variables to stabilize the division with weak denominators, where $k1$=0.01 and $k2$=0.02. In Eq. (11), Q is the maximal intensity value of x and y images, and MSE is the mean squared error. In general, for both SSIM and PSNR, higher

values indicate better prediction, while lower MAE suggests better agreement between synCT and CT-SIM data. Consideration was given to agreement using the full field of view, bone, air and soft tissue regions. To segment the air masks in the images, a threshold of -465 HU was set based on values obtained for the upper airway from the literature[34]. For bone regions, a threshold of +200 HU was implemented in both real and generated CTs. Comparisons were also made between CNN and GAN approaches to synCT generation.

Results

Experimental setup

The generator and discriminator networks were trained between MRI and CT image pairs, and weights were initialized from a Gaussian distribution with parameter values of 0 and 0.02 for mean and standard deviation, respectively. The model is trained using minibatch stochastic gradient descent and Adam solver [35] and batch size of 1. The GAN model stabilized after 200 epochs (i.e., one forward pass and one backward pass of all the training samples). Here, training was considered to be converged when the training error between two adjacent epochs was < 0.001.

GAN results

GAN training took approximately 11 hours with a GeForce 980 Ti GPU with testing taking 5.7±0.6 seconds. Table 1 best summarizes the MAE, PSNR and SSIM metrics computed based on the real and synthetic CTs for each fold in the five-fold validation.

Table 1 summarizes the three quantitative measurements on the full field of view (FOV) of the entire head for all fifteen cases. The average MAE across the entire FOV for all fifteen patients was 89.30±10.25 Hounsfield units (HU); the average PSNR was 26.64±1.17, and the average SSIM was 0.83±0.03. To better evaluate the performance of the proposed model, the MAEs for different regions of the brain, including bone, air and soft tissue regions are summarized in Table 2. After calculating the MAE in the other regions, we found the largest errors near the segmented bone and air regions (MAE=255.22±47.74 and 240.94±60.21 HU, respectively). However, in the remaining tissue regions the MAE was found to be lower, with values of 41.85±8.58 HU.

Impact of Least Squares objective function

The average full FOV MAEs obtained across the cohort were 100.3 HU and 89.3 HU for the proposed GAN model using the negative log likelihood and the least squares loss functions, respectively. Overall, synCTs obtained with the least squares loss function had a higher similarity to the real CTs with fewer artifacts than those generated with the negative log likelihood loss function (results not shown).

Impact of Adversarial Learning

When comparing our GAN model to a ResNet model with no discriminator block, the GAN model had higher spatial resolution (less blurry), particularly for edge features, and had a higher similarity to the corresponding CTs as shown in Figure 3. The average full FOV MAEs were 92.6 HU and 89.3 HU for the ResNet model and the proposed GAN model, respectively.

Comparison of CNN and GAN

The average MAE, PSNR and SSIM obtained for all fifteen patients with the full FOV using the CNN model is 102.41±11.13 HU, 25.43±1.09 and 0.79±0.03 respectively while the average MAE, PSNR and SSIM using our GAN model is 89.30±10.25 HU, 26.64±1.17 and 0.83±0.03 respectively. Figure 4 highlights four randomly selected synCT patient cases generated by GAN and CNN. Qualitative comparisons for two different methods show that the synCTs generated by the proposed GAN model are more similar to the real CTs, less noisy, and have more preserved details compared to CNN results. SynCTs generated by the CNN method show larger errors in the skull and bone areas and also in the boundaries in the residual images. The 2nd row highlights a case where the bone at the posterior of the skull was preserved with GAN but was eroded with CNN.

Figure 5 highlights a detailed comparison between two synCT images generated by GAN and CNN. The GAN output was found to be less diffuse and more accurate than the CNN model. The zoomed in regions illustrate that the GAN model synthesized regions near the bone/air interfaces more accurately than CNN in both examples. GAN also preserved smaller, more complex details such as the auditory canal, thinner bones, and the posterior of the skull.

Finally, P8 and P13 in Figure 6 had two pieces of skull removed during surgery that presented challenges for the proposed model. Figure 6 displays a comparison of several sagittal slices between real CT and synCT for these two cases, representing the worst-case scenarios. For these patients, the synCT rows of Figure 6 show that the frontal bone that was affected by surgery is well represented in the synCT images.

Discussion

In this work, we introduced a novel method for generating synCTs from T1-weighted post-Gd MRI datasets by applying a GAN model that uses ResNet architecture as the generator and regular CNN architecture as the discriminator. To quantify potential advantages of implementing GAN, we then compared our results to a trained CNN. Our experimental results show that the proposed GAN model can efficiently and accurately generate synCT images from the MRI input while outperforming CNN, thus offering strong potential for supporting MR-only radiation therapy workflows. The overall MAE of our approach across the entire head for the 15 patient cohort was 89.30±10.25 HU. Our results were slightly higher than those obtained by Han (84.8 ± 17.3 HU) for a CNN model, although the studies used different patient populations. In our comparison of CNN and GAN on the same cohort yielded slightly lower MAE for GAN than CNN. However, the largest advantage of using GAN compared to CNN is that GAN has been shown in this study (Figure 5) to preserve details and yield a better representation of the initial input image.

Recently, Nie et al. generated synthesized images using context-aware GAN using a 3D model [25] whereas the current work uses 2D datasets. Two-dimensional models require less training data[20], GPU memory, and computational time. One limitation in reference [25] is that due to limited training data, the model was trained on patches that may not provide sufficient image context information [20] and may affect the network's modeling capacity. Other differences include using an Auto-Context Model to iteratively train several GANs to overcome this problem, which leads to longer training time. By comparison, our GAN model automatically learns a mapping to generate a synCT from a full MRI slice. The two models use different loss functions (cross entropy vs. our selection of least squares loss) and reconstruction error functions (Nie et al. employed L2 loss and an image gradient difference loss function while we used L1).

We have illustrated that our selection of the least squares loss function yielded more realistic images with lower MAE, while L1 encourages less blurring synthetic CTs compared to using L2[24]. Least squares loss has been shown to be more stable during GAN training and generate higher quality results than other loss functions[28].

Our work suggests that incorporating adversarial learning generates more realistic synCTs with higher spatial resolution and lower MAE in the brain than CNN with no discriminator block. GAN has been shown to outperform CNN for image generation[24, 32, 33], which is consistent with our work. Our GAN technique generated synCT images with improved feature preservation and higher overall agreement.

Another advantage of the proposed GAN method is the short time that it requires for producing the synCT maps once the initial training has been implemented. Although the time required for training the GAN ~11 hours, once the network is trained, producing the synCT maps directly from the T1-Weighted MR images can be completed in <6 seconds for a variety of slice numbers. This feature makes the GAN method a practical tool that can be integrated into real-time applications for producing synCT maps from MR images. Recent work by Han derived synCTs of 9 seconds for 160 MR images using CNN[20]. Other synCT methodologies require complex processing pipelines such as unwrapping phase images [12] or combining multiple MRI datasets that require additional processing time and may introduce additional sources of uncertainty.

The GAN was trained using only post Gadolinium T1-Weighted images. To further study the feasibility of this method, training and testing can be performed using other contrasts such as pre-contrast images or other tissue weightings. T1-Weighted images were selected for training because they had the highest in- and through-plane resolution of all the images available for the

patient cohort and are widely used in brain cancer RTP. Other approaches require specialty sequences such as mDIXON, UTE, or PETRA sequences[9, 12, 19, 36]. To make this method more robust, MR images from scanners with different field strengths can be used in the training set.

It should be noted that one of the limitations of the current study is the relatively small sample size that was used for training, however this sample size is similar to what has been reported in the literature [20, 24]. To further improve the performance of this method, images from a larger number of patients can be incorporated into the training set in order to further optimize the training sample size for the current application. This may help improve the performance of the GAN in reducing errors in regions near the bone and air interfaces. Nevertheless, the robustness of our method was demonstrated in several worst-case scenarios such as areas where the skull was displaced due to surgery, where excellent agreement was observed even in these complex cases.

Although GAN models have shown great success on a variety of imaging tasks, their training process may be challenging. Potential issues in GAN training include vanishing gradient and mode collapse (i.e., where the generator produces the same outputs) that may lead to problematic convergence behavior. As shown in [28] and adopted in our work, the least squares objective function can help overcome the vanishing gradient problem in GAN training, resulting in a more stable model with smaller degree of mode collapse. More recently, Wasserstein GAN (WGAN) was recently proposed that can lead to more stable training[37] which can be explored in future synCT work. Future work may include performing a dosimetric analysis using original CT-SIM and synCT datasets and extending the GAN to other disease sites. However, areas such as the pelvis or abdomen may be more challenging due to differences in respiratory or

physiological states whereas the brain does not have a tendency to these types of effects with the exception of post-surgical changes.

Conclusion

In this paper, we introduced a GAN model to generate synCTs using T1-weighted MRI as the input. Our GAN model contains two competing architectures, i.e., ResNet and CNN, as the generator and discriminator, respectively. Experimental results show that the proposed method can achieve superior results on synCTs than CNN, offering strong potential for supporting MR-only workflows.

References

1. C. Njeh, "Tumor delineation: The weakest link in the search for accuracy in radiotherapy," Journal of medical physics/Association of Medical Physicists of India **33**, 136 (2008).
2. V. S. Khoo, and D. L. Joon, "New developments in MRI for target volume delineation in radiotherapy," The British journal of radiology **79 Spec No 1**, S2-15 (2006).
3. E. P. M. Jansen, L. G. H. Dewit, M. van Herk, and H. Bartelink, "Target volumes in radiotherapy for high-grade malignant glioma of the brain," Radiotherapy and Oncology **56**, 151-156 (2000).
4. G. A. Whitfield, S. R. Kennedy, I. K. Djoukhadar, and A. Jackson, "Imaging and target volume delineation in glioma," Clinical oncology (Royal College of Radiologists (Great Britain)) **26**, 364-376 (2014).
5. K. Ulin, M. M. Urie, and J. M. Cherlow, "Results of a multi-institutional benchmark test for cranial CT/MR image registration," International Journal of Radiation Oncology* Biology* Physics **77**, 1584-1589 (2010).
6. H. Nakazawa, Y. Mori, M. Komori, Y. Shibamoto, T. Tsugawa, T. Kobayashi, and C. Hashizume, "Validation of accuracy in image co-registration with computed tomography and magnetic resonance imaging in Gamma Knife radiosurgery," Journal of radiation research **55**, 924-933 (2014).
7. G. Opposits, S. A. Kis, L. Trón, E. Berényi, E. Takács, J. G. Dobai, L. Bognár, B. Szűcs, and M. Emri, "Population based ranking of frameless CT-MRI registration methods," Zeitschrift für Medizinische Physik **25**, 353-367 (2015).
8. B. Demol, C. Boydev, J. Korhonen, and N. Reynaert, "Dosimetric characterization of MRI-only treatment planning for brain tumors in atlas-based pseudo-CT images generated from standard T1-weighted MR images," Medical Physics **43**, 6557-6568 (2016).
9. Y. Yang, M. Cao, T. Kaprealian, K. Sheng, Y. Gao, F. Han, C. Gomez, A. Santhanam, S. Tenn, and N. Agazaryan, "Accuracy of UTE-MRI-based patient setup for brain cancer radiation therapy," Medical physics **43**, 262-267 (2016).

10. J. H. Jonsson, M. M. Akhtari, M. G. Karlsson, A. Johansson, T. Asklund, and T. Nyholm, "Accuracy of inverse treatment planning on substitute CT images derived from MR data for brain lesions," Radiation Oncology **10**, 1 (2015).
11. R. G. Price, J. P. Kim, W. Zheng, I. J. Chetty, and C. Glide-Hurst, "Image guided radiation therapy using synthetic computed tomography images in brain cancer," International Journal of Radiation Oncology* Biology* Physics **95**, 1281-1289 (2016).
12. W. Zheng, J. P. Kim, M. Kadbi, B. Movsas, I. J. Chetty, and C. K. Glide-Hurst, "Magnetic Resonance–Based Automatic Air Segmentation for Generation of Synthetic Computed Tomography Scans in the Head Region," International Journal of Radiation Oncology* Biology* Physics **93**, 497-506 (2015).
13. S.-H. Hsu, Y. Cao, T. S. Lawrence, C. Tsien, M. Feng, D. M. Grodzki, and J. M. Balter, "Quantitative characterizations of ultrashort echo (UTE) images for supporting air–bone separation in the head," Phys. Med. Biol. **60**, 2869–2880 (2015).
14. E. Paradis, Y. Cao, T. S. Lawrence, C. Tsien, M. Feng, K. Vineberg, and J. M. Balter, "Assessing the Dosimetric Accuracy of Magnetic Resonance-Generated Synthetic CT Images for Focal Brain VMAT Radiation Therapy," International Journal of Radiation Oncology*Biology*Physics **93**, 1154-1161 (2015).
15. E. R. Kops, and H. Herzog, "Alternative methods for attenuation correction for PET images in MR-PET scanners," in *Nuclear Science Symposium Conference Record, 2007. NSS'07. IEEE*(IEEE2007), pp. 4327-4330.
16. M. Hofmann, F. Steinke, V. Scheel, G. Charpiat, J. Farquhar, P. Aschoff, M. Brady, B. Schölkopf, and B. J. Pichler, "MRI-based attenuation correction for PET/MRI: a novel approach combining pattern recognition and atlas registration," Journal of nuclear medicine **49**, 1875-1883 (2008).
17. N. Burgos, M. J. Cardoso, F. Guerreiro, C. Veiga, M. Modat, J. McClelland, A.-C. Knopf, S. Punwani, D. Atkinson, and S. R. Arridge, "Robust CT synthesis for radiotherapy planning: application to the head and neck region," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*(Springer2015), pp. 476-484.
18. P. D. Gatehouse, and G. M. Bydder, "Magnetic resonance imaging of short T2 components in tissue," Clin Radiol. **58**, 1-19 (2003).
19. A. Johansson, M. Karlsson, and T. Nyholm, "CT substitute derived from MRI sequences with ultrashort echo time," Medical physics **38**, 2708-2714 (2011).
20. X. Han, "MR-based synthetic CT generation using a deep convolutional neural network method," Medical Physics **44**, 1408-1419 (2017).
21. K. Simonyan, and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556 (2014).
22. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*(2016), pp. 770-778.
23. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*(2014), pp. 2672-2680.
24. P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," arXiv preprint arXiv:1611.07004 (2016).
25. D. Nie, R. Trullo, J. Lian, C. Petitjean, S. Ruan, Q. Wang, and D. Shen, "USEMedical image synthesis with context-aware generative adversarial networks," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*(Springer2017), pp. 417-425.
26. S. Ioffe, and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International conference on machine learning*(2015), pp. 448-456.
27. X. Mao, Q. Li, H. Xie, R. Y. Lau, and Z. Wang, "Multi-class Generative Adversarial Networks with the L2 Loss Function," arXiv preprint arXiv:1611.04076 (2016).

28. X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang, and S. P. Smolley, "Least squares generative adversarial networks," in *2017 IEEE International Conference on Computer Vision (ICCV)*(IEEE2017), pp. 2813-2821.
29. F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," IEEE transactions on medical imaging **16**, 187-198 (1997).
30. C. Studholme, D. L. G. Hill, and D. J. Hawkes, "An overlap invariant entropy measure of 3D medical image alignment," Pattern Recognition **32**, 16 (1999).
31. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*(Springer2015), pp. 234-241.
32. X. Huang, and S. Belongie, "Arbitrary Style Transfer in Real-time with Adaptive Instance Normalization," arXiv preprint arXiv:1703.06868 (2017).
33. C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, and Z. Wang, "Photo-realistic single image super-resolution using a generative adversarial network," arXiv preprint arXiv:1609.04802 (2016).
34. H. Nakano, K. Mishima, Y. Ueda, A. Matsushita, H. Suga, Y. Miyawaki, T. Mano, Y. Mori, and Y. Ueyama, "A new method for determining the optimal CT threshold for extracting the upper airway," Dentomaxillofac Radiol **42**, 26397438 (2013).
35. D. Kingma, and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980 (2014).
36. S.-H. Hsu, Y. Cao, T. S. Lawrence, C. Tsien, M. Feng, D. M. Grodzki, and J. M. Balter, "Quantitative characterizations of ultrashort echo (UTE) images for supporting air–bone separation in the head," Physics in medicine and biology **60**, 2869 (2015).
37. M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein gan," arXiv preprint arXiv:1701.07875 (2017).

Figure 1. The architecture of GAN model used for generating SynCTs.

Figure 2. ResNet architecture highlighting the input MRI and output synthetic CT. Shortcut connections are shown as solid arrows.

Figure 3. Qualitative comparison of synCTs generated by the proposed GAN model and a CNN model (ResNet). The selected areas on each image are enlarged in its next column to highlight enhanced details retained by GAN.

Figure 4. Qualitative comparison of synCTs generated with GAN and CNN (U-net) for four different patients. The input T1-weighted Post-Gadolinium MRI, the corresponding real CT, the synCT generated by GAN, the residual image by GAN (subtracting the real CT from the synCT), the synCT generated by CNN, the residual image by CNN are shown.

Figure 5. Qualitative comparison of synCT generated by CNN (U-net) and the proposed GAN model. First column: real CT, second column: synCT generated by our GAN model, third column: synCT generated by CNN. The selected areas on each image are enlarged below that image.

Figure 6. Qualitative comparison of the GAN prediction result for Patient 8 and 13 presented the worst-case scenarios, highlighting close agreement of bone, air, and tissue across several sagittal slices despite the presence of postsurgical regions in the skull.