A feature invariant generative adversarial network for head and neck MRI/CT image synthesis

Article *in* Physics in Medicine and Biology · March 2021

DOI: 10.1088/1361-6560/abf1bb

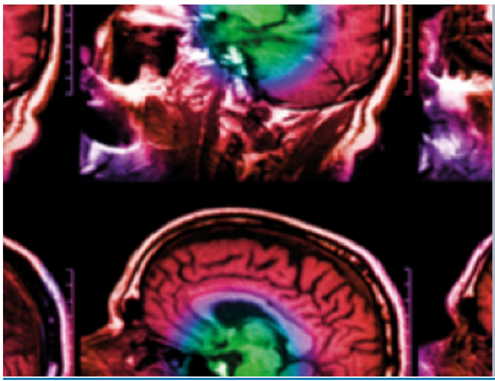
Physics in Medicine & Biology

IPEM
Institute of Physics and
Engineering in Medicine

PAPER

A feature invariant generative adversarial network for head and neck MRI/CT image synthesis

To cite this article: Redha Touati *et al* 2021 *Phys. Med. Biol.* **66** 095001



Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

RECEIVED
7 December 2020

REVISED
19 March 2021

ACCEPTED FOR PUBLICATION
24 March 2021

PUBLISHED
23 April 2021

PAPER

A feature invariant generative adversarial network for head and neck MRI/CT image synthesis

Redha Touati[1,*], **William Trung Le**[1] **and Samuel Kadoury**[1,2]

[1] MedICAL Laboratory, Polytechnique Montreal, Montreal, QC, Canada

[2] CHUM Research Center, Montreal, QC, Canada

[*] Author to whom any correspondence should be addressed.

E-mail: redha.touati@polymtl.ca, william.le@polymtl.ca **and** samuel.kadoury@polymtl.ca

Keywords: MRI-CT image synthesis, conditional GAN, feature invariant learning, multi-resolution edge features, histogram matching, head and neck radiotherapy

Abstract

With the emergence of online MRI radiotherapy treatments, MR-based workflows have increased in importance in the clinical workflow. However proper dose planning still requires CT images to calculate dose attenuation due to bony structures. In this paper, we present a novel deep image synthesis model that generates in an unsupervised manner CT images from diagnostic MRI for radiotherapy planning. The proposed model based on a generative adversarial network (GAN) consists of learning a new invariant representation to generate synthetic CT (sCT) images based on high frequency and appearance patterns. This new representation encodes each convolutional feature map of the convolutional GAN discriminator, leading the training of the proposed model to be particularly robust in terms of image synthesis quality. Our model includes an analysis of common histogram features in the training process, thus reinforcing the generator such that the output sCT image exhibits a histogram matching that of the ground-truth CT. This CT-matched histogram is embedded then in a multi-resolution framework by assessing the evaluation over all layers of the discriminator network, which then allows the model to robustly classify the output synthetic image. Experiments were conducted on head and neck images of 56 cancer patients with a wide range of shape sizes and spatial image resolutions. The obtained results confirm the efficiency of the proposed model compared to other generative models, where the mean absolute error yielded by our model was 26.44(0.62), with a Hounsfield unit error of 45.3(1.87), and an overall Dice coefficient of 0.74(0.05), demonstrating the potential of the synthesis model for radiotherapy planning applications.

1. Introduction

Within radiotherapy workflows, computed tomography (CT) and magnetic resonance imaging (MRI) modalities have a wide range of clinical applications such as tumor volume localization, clinical pathology assessment, radiotherapy treatment planning, and registration for image guidance systems, all of which benefit from multi-modal planning methods (Li *et al* 2019, 2019, Liang *et al* 2019, Liu *et al* 2019, Kazemifar *et al* 2020, Oulbacha and Kadoury 2020). Indeed, both modalities are most frequently used together as they provide complementary information, which significantly improves medical diagnosis and subsequent treatment of diseases. For external beam radiotherapy procedures, the treatment is often dependent on the location and stage of the cancer. These are usually revealed by the diagnostic MRI, identifying the tumors at an early stage. The therapeutic objective then consists of balancing treatment efficacy while minimizing toxicity through restriction of the delivered dose and affected area. MRI in particular provides improved tumor segmentation, organs at risk, and patient positioning, due to its superior soft tissue contrast. Its limitations however arrives at the treatment planning stage: it does not provide the necessary electron density information used for dose calculation. In contrast, CT images provide the necessary dose absorption information per tissue based on Hounsfield units

(HU), which is required for dose delivery calculations. Availability of both sequences at time of treatment is not always assured, due to several factors including limited hospital resources and time constraints of the patient, and additional radiation in the case of CT imaging. For these reasons, MR-based workflow is desirable, as it would provide both its high quality imaging properties and necessary dose attenuation levels for radiotherapy treatment planning.

In medical imaging, multi-modality image synthesis is defined as a transformation process that aims to create new realistic images indistinguishable from the original image representation, based on an image of the same anatomy but originating from another modality (Frangi *et al* 2018). This process is particularly challenging when seeking to generate a mapping from high-resolution images with the additional presence of abnormalities such as tumors. In this case, the synthesis process consists of transforming voxel intensities from one imaging modality to another. The images can be obtained from completely different acquisition systems—such as CT and MRI—based on entirely distinct physical principles. Indeed, the transformations need to be performed under the assumptions that the images will not share the same anatomical characteristics (bone vs soft tissue) as well as possess highly heterogeneous appearances for the considered patient.

Several previous works have been proposed in the literature for generating automatically synthetic CT (sCT) images from MRI scans. Initially introduced for medical attenuation correction purpose in positron emission tomography (PET) (Hofmann *et al* 2008), conventional medical image synthesis methods can be generally divided in three categories: density, single/multiple atlas-based and learning based techniques. The most popular methods consists of density-based techniques using pre-defined thresholds and morphological operations for the delineation of volumes of interest according to the tissue class labels. These then assign a density value (electronic or physical) to each class region defining the sCT image (, Keereman *et al* 2010, Rank *et al* 2013,). The major limitation of these approaches is the need of manual intervention for tuning the relative intensity thresholds and choosing the appropriate bulk density values. The category of atlas-based methods consists of performing a registration between one or several atlases of MRI scans, with corresponding CT images and eventually producing an organ label map, with which the input patient MRI scan can be converted to (Stanescu *et al* 2008, Johansson *et al* 2013). In multi-atlas techniques, the sCT image is generated by merging multiple atlases together after an alignment with the patient MRI (Prabhakar *et al* 2007, Jonsson *et al* 2013, Korsholm *et al* 2014). These methods are however more sensitive to atypical patient anatomies due to the fact that they are based on the structural similarity between the patient and the atlases (Johansson *et al* 2013). They also still require manual tuning of the optimal number of atlases to be used in the registration process (Kazemifar *et al* 2020) which may cause inter-patient registration errors. Finally, traditional machine learning methods consist of capturing the relationship between intensities of MRI and CT voxels in the HU space based on a learned criterion such as a regression model or a combination of classification and general density assignment (Robson *et al* 2003, Van der Bom *et al* 2011, Dowling *et al* 2012, Metcalfe *et al* 2013, Huynh *et al* 2015). In this category, we can also find hybrid methods that combine machine learning with atlas-based techniques based methods, which depend on manually selected features (Kazemifar *et al* 2020).

To overcome these issues, recently proposed approaches make use of deep learning techniques for addressing several challenges in image diagnosis (Işın *et al* 2016, Litjens *et al* 2017). More specifically, generative models based on adversarial networks (GAN) (Goodfellow *et al* 2014) are particularly well suited for domain translation problems. Such methods for image synthesis include the pseudo 3D Cycle GAN (Oulbacha and Kadoury 2020) proposing to synthesize a CT image from MRI spine image for image-guided applications, using a 3D fully convolutional network (FCN) to synthesize a CT image from the pelvic area in the MRI (Nie *et al* 2016), as well as the DualGAN (Yi *et al* 2017) and Cycle GAN (Zhu *et al* 2017), models designed for CT synthesis of the brain (Han 2017) and in the pelvic area (Dong *et al* 2017) respectively. More recently, conditional GAN networks have shown impressive results to improve the quality of the sCT image estimation (Wolterink *et al* 2017, Nie *et al* 2017) with the requirement that paired images be available for the input and target modalities, precisely the GANs models use U-net networks to deal with the anatomy challenges (Klages *et al* 2020, Liu *et al* 2020, Qi *et al* 2020).

Unlike previous deep learning synthesis models, in this work we are particularly interested in estimating sCT images from MRI acquisitions of the head and neck region, which present significant challenges due to the wide variety in patient anatomies and due to the different spatial resolutions. Our proposed multi-modal synthesis model is designed with a conditional GAN architecture, in which we first propose a new invariant imaging feature representation to match the common structural regions of the generated and the original CT images in terms of high-frequency and appearance components, which improves robustness of the generator towards variations in anatomical boundaries. This representation is based on a convolutional discriminator in the latent space created by the GAN, encoded with compact and complementary features. Second, an evaluation of the sCT image is assessed at different resolution scales within the discriminator, helping to guide the training of the final classifier.

The remainder of the paper is structured as follows. Section 2 presents the proposed deep CT-synthesis model and its feature invariant learning representation. Section 3 presents the different evaluation strategies using two distinct clinical datasets, with the qualitative and quantitative synthesis results presented on both sets, and provides a comparison with the state-of-the-art deep learning techniques. Section 4 discusses the results and the experimental limitations, while Section 5 concludes the paper.

2. Materials and methods

In this work, we treat pairs of images acquired for the same patient, which are obtained from two different modalities (MRI and CT) in the patient's head and neck region, providing MRI/CT pair image volumes. We also used image pairs which already rigidly co-registered between the CT and MRI. The proposed CT-synthesis model consists of learning a suitable representation that helps to address the different issues occurring from the diversity of anatomies and levels of contrast in images. The model is designed for medical image synthesis problems, handling with the anatomical variability challenge.

2.1. Clinical dataset description

We use a first dataset (DS1) for training and testing purposes, consisting of 36 real MRI/CT pairs head and neck imaging volumes coming from 36 patients, scanned between 2015 and 2017. These different imaging sequences are provided by the department of radiation oncology at the Center Hospitalier de l'Université de Montréal and includes MRI T1-weighted images, as well as CT volumes. The paired data are divided after being pre-processed into a set containing MRI T1-weighted images using Gradient Echo sequence type (GR) with segmented k-space (SK) or MAG prepared (MP) sequence variants, as well as CT volumes using 120–140 kVp energy levels, and a TR of 4000 ms and TE of 12ms. The imaging data was acquired from different subjects that reflect a variety of head and neck shapes with different sizes and under different spatial resolution images (varied from 0.68 to 0.98mm), and acquired with 2 different 3T clinical scanners (Siemens Magnetom, Philips Achieva) used for general diagnosis. From these 36 patients, 25 had acquisitions using the Philips Achieva scanner and 11 using the Siemens Magnetom scanner. A second dataset (DS2) consists of sagital slices and was composed of 20 different patients was used as a hold-out testing set, with an in plane resolution of 0.8mm from a 3T Siemens Magnetom recruited in 2019. All MR images were rigidly registered to the reference CT by a radiation-oncology specialist and each CT/MRI-T1 volume comprised seventy (70) 2D axial image slices with size equal to 256×256. The CT intensity values are encoded in the HU space. Images were pre-processed, which included background removal such as tables and noise reduction.

2.2. Feature invariant learning representation

Representation learning relies on the selection of relevant features, without considering the non-pertinent and redundant features for solving the task at hand (Zhong *et al* 2016). To this end, the trained CT-synthesis model is mainly based on the combination of complementary representations (see Figure 1), which helps to improve the efficiency of the learning process in the CT-synthesis application. More specifically, learning from these representations aims at finding the common structural information in terms of high-frequency patterns between similar and dissimilar regions of the generated sCT and the real CT images which exhibit salient features, ideally identifying and placing more importance on the same high-level features. Our goal is to reinforce the learning of the model to synthesize a sCT image that shares the same high-frequency feature distribution to the considered region within the real CT image. High-frequency patterns such as contours are relatively invariant features between two areas from paired images, representing the same anatomical object. This helps to delimit regions and provides valuable information about the internal region's shape orientation that can be used as a high edge feature to guide the learning of the proposed model in order to improve the CT-synthesis quality map. Nevertheless, relying solely on edge feature representations may not be as reliable as using information about the intensity distribution and the appearance of the entire image as well as their internal regions.

A straightforward way to improve the representation learning process consists of integrating local intensity distributions, which can be used as local appearance information. This way, combining both edge and appearance features leads to a feature vector representation with complementary characteristics of CT/MR image patterns, especially in a deep representation framework, producing additional discriminant features that can be embedded in the latent space. To achieve this, we rely on a convolutional neural network (CNN) backbone architecture, in which we propose to encode the deep feature space resulting from the CNN network by quantized histograms. These histograms integrate multi-level features reflecting the different network layers, as shown in Figure 1. We first compute the quantized intensity histogram that represents the intensity distribution, followed by the quantized histograms of edge attributes describing the intensity variations across

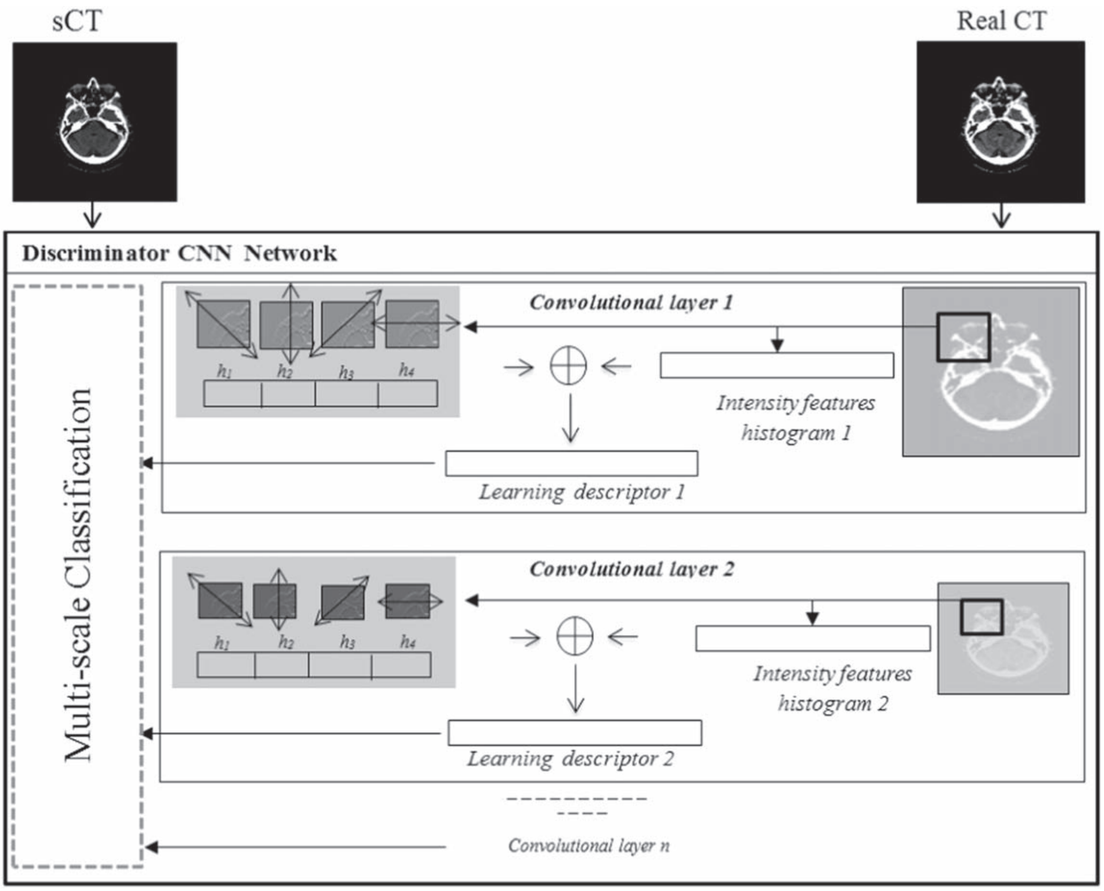


Figure 1. The proposed multiscale GAN discriminator-based feature invariant representation. The learning representation models each convolutional map by a combination of intensity and edge orientations histograms, defined for each squared window and for each layer. Here, h_1, h_2, h_3, h_4 are respectively oriented quantized histograms for the left diagonal, the vertical, the right diagonal and the horizontal directions with equidistant binning. The multiscale classifier considers this hierarchical representation as its inputs.

the CT patterns (Dalal and Triggs 2005, Gonzalez and Woods 2018). The magnitude and orientation feature values are computed in four different directions, the right and left diagonals as well as the horizontal and the vertical directions. The gradient magnitude J of a point p in the direction p_x can therefore be defined as:

$$\nabla J = (J_{p_x}) = \left(\frac{\delta J}{\delta p_x}\right) \tag{1}$$

$$||\nabla J|| = |J_{p_x}|. \tag{2}$$

All of the quantized feature histograms are computed from a cubic window of fixed size 8×8, which contains a set of intensities in a neighborhood centered by the intensity value being characterized. Each layer is encoded by histograms of edge orientations combined with the intensity histogram. The quantized histograms have the same number of bins, regardless of the levels within the CNN network. Consequently, the resultant representation characterizes the CT-image inter- and intra-pattern variations through intensity and edge attributes extracted from the CNN feature space.

2.3. Multimodal Synthesis Model

From a network architecture perspective, the CT-synthesis model architecture is a closely linked variant to the generative adversarial network (GAN) model, as well as the so-called conditional GAN (Isola *et al* 2017), which is well adapted to the medical image synthesis problem. Let us recall that in the GAN architecture, the model consists of two CNNs. The first CNN is a generator G that produces an image candidate, while the second network acts as a discriminator D, comparing the generated image candidate with the real reference image to classify the candidate image in the 'real' or 'fake' class. The learning process is repeated until the discriminator can no longer make the distinction between the reference and the synthetic image. In our work, the overall CT-synthesis model makes use of the conditional GAN architecture, using a similar adversarial loss function to build the model (Isola *et al* 2017). The loss is optimized with the mixed min-max objective, combining the global $\zeta_{cGAN}(G, D)$ loss with an additional L_1 loss term to learn a mapping from observed image x and random noise vector z, to output image y. The final combined cost function is given in equation (3):

$$G^{*} = \text{argmin}_{G} \max_{D} \zeta_{cGAN}(G, D) + \lambda \zeta_{L_1}(G), \tag{3}$$

where the objective function for the generator $\zeta_{cGAN}(G, D)$ and the $\zeta_{L_1}(G)$ terms are given respectively by equations (4) and (5), with λ as the $\zeta_{L_1}(G)$ loss weight:

$$\zeta_{cGAN}(G, D) = \mathbb{E}_{x,y}[\log D(x, y)] + \mathbb{E}_{x,z}[\log(1 - D(x, G(x, z)))] \tag{4}$$

$$\zeta_{L_1}(G) = (\mathbb{E}_{x,y,z}) \, \|y - G(x, z)\|_1. \tag{5}$$

In the min-max equations (3)–(5) which are based on the expected cross-entropy loss, the symbols $G(z)$ and $D(y)$ are the generator and the discriminator networks, respectively. The variables x, z and y are the observed conditioned (MRI) image, the noise vector and the output image (CT) extracted from the quantized histograms, respectively. The CT image is distributed following an unknown probability distribution p_y. The generator G transforms the noise vector z into an image according the probability distribution p_y. The generated image is then classified with the discriminator D that uses our proposed feature representation based on the learned feature space. The discriminator D and the generator G are trained alternately so that the discriminator D attempts to maximize the expected objective function while G attempts to minimize it.

Here, we reproduce both the U-net convolutional network for the generator as well as the convolutional PatchGAN structure for the discriminator (Isola *et al* 2017). The training of the model is achieved using the feature representation derived from the PatchGAN feature space (see Section 2.2).

The proposed feature representation guides the generator in learning how to predict a new sCT mapping that proportionally preserves the multiscale edges and the appearance features of the real CT image. These features are extracted from each convolutional feature map and used in the discriminator to output a 'real' or 'fake' classification. First, the generator synthesis model transforms the input MRI space into a new sCT space using the U-net module (Ronneberger *et al* 2015). Then, in the discrimination phase, the transformed sCT and the ground truth CT images are passed through the convolutional PatchGAN layers, and the resulting deep feature space is encoded by the proposed appearance and edge features representation (see Figures 1 and 2). In more details, the CNN network maps the CT and sCT images into multiscale convolutional features maps through multiple layers. We model each convolutional layer by an encoded descriptor representing quantized histograms that captures the intensity distribution, and the four different edge feature orientations. These high level complementary histograms have equidistant binning in all existing levels in the PatchGAN network. Once the descriptors are constructed for each CNN layer, the synthetic sCT and the CT descriptors are evaluated and compared. The discriminator classifier considers a hierarchical framework based on multiresolution histogram representations of the convolutional layers maps. Considering that the representation of each convolutional maps are distributed in the vertical and horizontal directions by a factor of 2 at each scale, the discriminator output will consider a set of image feature representations encoding the images in a set of details that appear not only in the given specific resolution level, but also at different spatial resolution scales, allowing to integrate more relevant information. Therefore, the comparison between the real CT and the sCT histograms is done by considering all scales, so that the averaging score of all scales is assessed to decide if the requested synthetic sCT-image is *real* or *fake* (see Figures 1 and 2).

2.4. Network architecture details

In the conditional GAN with a hierarchical feature combination, the structure of the CNN discriminator consists of a convolutional classifier with 5 layers. The network applies four convolutional filters of size 4×4, with a stride of 2 and a padding of 1. The last layer is also a convolutional layer which uses a convolutional filter of size 4×4 with zero padding and a stride of 1, followed by a sigmoid function. Dropout and LeakyRelu operations are applied to the first 4 layers, while an Instance Norm operation is used for the second, third and fourth layers (see Table 1), using a scale of 0.2 and a dropout rate of 0.5. The generator is based on a U-net structure and consists of an encoder and a corresponding CNN decoder with symmetrical skip connections. The network depth is set at (5) levels. At each level of the encoder, two 3×3 convolutions and a Relu activation function are performed for each convolutional layer, with a one max-pooling operation that progressively down-samples the input with a stride of 2. This process is repeated until the bottleneck layer is reached, and at each layer, the number of feature maps is doubled. In the fifth layer, two 3×3 convolutional operations are used followed by a ReLU activation function. In the decoding phase, the decoder estimates the image from the encoded representation, in which we repeat the same process as in the encoder phase, but by applying a 2×2 up-sampling transpose convolutional operation, concatenation, two 3×3 convolutional operations with ReLU functions, which are applied at each level. The encoded image size is expanding and the number of features is halved at each level. The concatenation is applied over the layers to combine the encoding with the

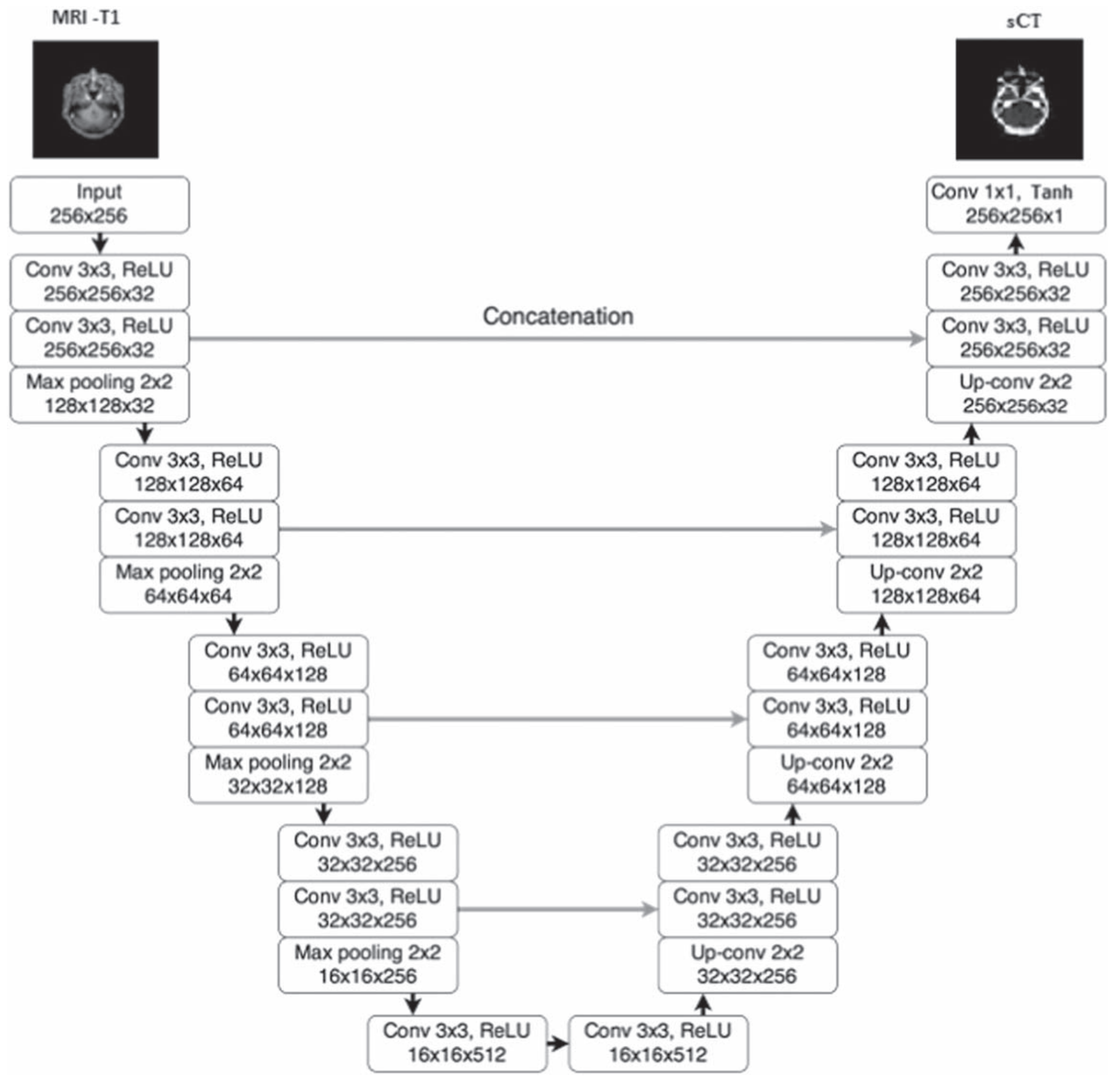


Figure 2. Symmetric 5-layer GAN generator based on a U-Net network, including skip connections between the encoder and the decoder block layers.

Table 1. Structure of the discriminator with details of the architecture with 5 convolutional layers, using a size filter of 4 × 4.

Name	Type	Filter number	Filter Size conv	Pad	Stride
Layer 1	Convolution/Dropout/LeakyRelu	64	4 × 4	1	2
Layer 2	Convolution/InstanceNorm/Dropout/LeakyRelu	128	4 × 4	1	2
Layer 3	Convolution/InstanceNorm/Dropout/LeakyRelu	256	4 × 4	1	2
Layer 4	Convolution/InstanceNorm/Dropout/LeakyRelu	512	4 × 4	1	2
Layer 5	Convolution/Sigmoid	1	4 × 4	0	1

corresponding decoded feature maps. The last layer acts as the output layer consisting of one convolutional layer of one feature map with a filter size of 1 × 1, followed by *Tanh* activation. Tables 1 and 2 summarize the architecture of the discriminator and the generator networks used in our proposed model.

2.5. Implementation details and experimental setting

For all experiments, the evaluation of the models was achieved using a cross-validation technique for DS1, and a holdout evaluation technique for DS2. First, a 6-fold cross-validation training/testing procedure was performed to compare the obtained CT-synthesis images with other synthesis model results (Oulbacha and Kadoury 2020). The trained model reserves 6 volumes from the thirty-six 36 volumes for the testing phase, while the training was performed on the remaining 30 volumes. The training process is repeated six times for each fold. Second, we evaluate the trained models on a separate holdout clinical dataset (DS2) consisting of 20 unseen patient cases.

Table 2. Structure of the U-net generator network architecture with details of the convolutional and deconvolutional block layers. The number of filters is doubled after each encoding layer and halved at each decoding layer using the 3 × 3 convolutional filter sizes and 2 × 2 up-convolutional filter sizes.

Name	Type	Filter number	Filter size conv	Filter size pool	Filter size up-conv
Encoding Layer 1	Convolution/ReLU/Convolution/ReLU/Max-pooling	32	3 × 3	2 × 2	N/A
Encoding Layer 2	Convolution/ReLU/Convolution/ReLU/Max-pooling	64	3 × 3	2 × 2	N/A
Encoding Layer 3	Convolution/ReLU/Convolution/ReLU/Max-pooling	128	3 × 3	2 × 2	N/A
Encoding Layer 4	Convolution/ReLU/Convolution/ReLU/Max-pooling	256	3 × 3	2 × 2	N/A
Encoding Layer 5	Convolution/ReLU/Convolution/ReLU	512	3 × 3	N/A	N/A
Decoding Layer 5	Up-convolution/Convolution/ReLU/Convolution/ReLU	256	3 × 3	N/A	2 × 2
Decoding Layer 4	Up-convolution/Convolution/ReLU/Convolution/ReLU	128	3 × 3	N/A	2 × 2
Decoding Layer 3	Up-convolution/Convolution/ReLU/Convolution/ReLU	64	3 × 3	N/A	2 × 2
Decoding Layer 2	Up-convolution/Convolution/ReLU/Convolution/ReLU	32	3 × 3	N/A	2 × 2
Decoding Layer 1	Convolution/Tanh	1	1 × 1	N/A	N/A

During training, data augmentation was used by generating a random crop of the original image sizes and a random flipping image (Oulbacha and Kadoury 2020), yielding 1024 volumes. The model optimization is realized using the Adam optimizer (Kingma 2015) with a momentum of 0.5 and an initial learning rate of 0.0002 which is linearly reduced to 0 starting at epoch 200. The number of epochs is set to 600, with the L_1 loss weight $\lambda = 100$, and a number of bins of 32 for the histogram feature extraction. The batch size is fixed to 1.

2.6. Evaluation metrics

For the quantitative evaluation, the quality of the synthesized CT images compared to the reference CT is evaluated using different imaging metrics (Bae and Kim 2015, Huynh *et al* 2016, Dong *et al* 2017, Lauritzen *et al* 2019, Oulbacha and Kadoury 2020). We calculate the mean absolute error (MAE) and peak-signal-to-noise-ratio (PSNR), between intensity values of the ground truth CT image and the sCT image in the HU intensity space. We also calculate the mean structural similarity (MSSIM) index, which does not consider the intensity difference, but allows us to evaluate the contrast level and the anatomical structure of the sCT image compared with the corresponding ground truth (Dong *et al* 2017). The last metric measures the Pearson cross-correlation (PCC) coefficient (Lauritzen *et al* 2019). Besides image quality metrics, we also compute the Frechet inception distance (FID) (Heusel *et al* 2017), which captures the similarity between the generated sCT and the real CT, based on statistics of the compared sCT images to statistics of the real CT images. We also compare the three models using the sliced Wasserstein distance (SWD) (Deshpande *et al* 2018) that measures the overall deviation between the synthetic and real images. We also evaluate the Bhattacharyya distance (BD) (Kailath 1967) between the synthetic sCT and real CT histograms to measure the overlap of pixel intensity distributions. Finally, we report the Dice score (Crum *et al* 2006, Milletari *et al* 2016), which is determined within three areas of interest: bony structures, soft tissue and withing air regions. To do so, thresholding using known HU ranges (see Table 5) was performed to obtain anatomical segmentation of these regions of interest which could then be evaluated 6 using:

$$\text{Dice} = \frac{2 \times TP}{2 \times TP + FN + FP}, \tag{6}$$

where TP, FN, FP are respectively the true positives, the false negatives and the false positives. These measures are then used to compare the synthesis results.

In our application, we have chosen to use several imaging metrics as they are frequently used in the literature and state-of-the-art synthesis methods (Bae and Kim 2015, Huynh *et al* 2016, Dong *et al* 2017, Lauritzen *et al* 2019, Oulbacha and Kadoury 2020) to evaluate the quality of the generated image, as they offer complementary information from one another. The usage of the different criterion provide not only more information on the generated sCT image, but also allows us to better understand the behavior of the different models. Specifically, while MAE and PSNR metrics are direct error and image quality measures computed at a pixel-wise level (Huynh *et al* 2016, Dong *et al* 2017, Oulbacha and Kadoury 2020), MSSIM, FID and SWD are used to evaluate the contrast and the anatomy of the sCT image, which are important to assess tumor burden in medical imaging. PCC allows us to evaluate how similar the generated anatomies are at a global image-level (Lauritzen *et al* 2019). BD is an interesting metric for its ability to compare histograms, which in the case of CT images have meaningful HU values, important for downstream dose calculation tasks. Finally, due to the importance of isolating specific regions, bone, air and soft-tissue overlap segments were evaluated using the intersection shape Dice score.

3. Experimental Results

In order to assess the model's performance, and to evaluate the capacity of the proposed CT-synthesis framework to synthesize a new CT image (sCT) from a wide variety of diagnostic MRI head and neck acquisitions, we performed a series of experiments on the imaging datasets from a tertiary clinical center (Center Hospitalier de l'Université de Montréal). We compare the proposed model with two state-of-the-art image synthesis models: the Cycle GAN with a ResNet (He *et al* 2016) as the generator and the 2D conditional GAN with a U-Net (Ronneberger *et al* 2015) as the generator. Furthermore, both models use a PatchGAN architecture for the discriminator part of the network (Oulbacha and Kadoury 2020). We apply the same testing conditions as given in (Oulbacha and Kadoury 2020), i.e. using the optimized set of parameter values and under the same training/validation procedure. We evaluate the performance of the models using cross-validation and holdout techniques performed on two separate head and neck MR/CT imaging datasets. Furthermore, we evaluate the ability of the synthesis models to generate the synthetic MRI-T1 mapping from the CT image domain.

Table 3. MAE, PSNR, SSIM, and PCC evaluations of CT-synthesis of the proposed model and the state-of-the-art deep synthesis models, using the DS1 head and neck imaging dataset. (⇓) Lower values indicate improved image synthesis quality. (⇑) Higher values indicate improved image synthesis quality.

	MAE (std) HU (⇓)	**PSNR (std) DB (⇑)**	**MSSIM index (std) (⇑)**	**PCC (std) (⇑)**
Cycle GAN	70.14 (0.47)	26.08 (1.71)	0.82 (0.04)	0.88 (0.05)
Conditional GAN	40.23 (1.02)	32.31 (2.91)	0.90 (0.09)	0.94 (0.07)
Proposed model	26.44 (0.62)	36.97 (2.67)	0.93 (0.08)	0.98 (0.031)

Table 4. FID, SWD, and BD evaluations of CT-synthesis of the proposed model and the state-of-the-art deep synthesis models, using the DS1 head and neck imaging dataset. (⇓) Lower values indicate improved image synthesis quality. (⇑) Higher values indicate improved image synthesis quality.

	FID (std) (⇓)	**SWD (std) (⇓)**	**BD (std) (⇑)**
Cycle GAN	254.45 (0.15)	14.84 (0.06)	0.84 (0.03)
Conditional GAN	170.72 (0.11)	12.03 (0.08)	0.87 (0.02)
Proposed model	134.45 (0.21)	7.38 (0.01)	0.94 (0.05)

Table 5. Dice score of CT-synthesis on the DS1 head and neck dataset obtained by our model and the state-of-the-art synthesis models, through the three segmented areas. (⇑) higher is the value better is the quality synthesis.

	Dice score (std) (⇑)	**Bone area HU**
Cycle GAN	0.61 (0.06)	
Conditional GAN	0.65 (0.08)	(>300)
Proposed model	**0.74 (0.05)**	
	Dice score (std) (⇑)	**Air area HU**
Cycle GAN	0.78 (0.03)	
Conditional GAN	0.80 (0.03)	(< -100)
Proposed model	**0.82 (0.02)**	
	Dice score (std) (⇑)	**Soft-tissue area HU**
Cycle GAN	0.65 (0.07)	
Conditional GAN	0.69 (0.04)	($-100 >$ and <300)
Proposed model	**0.76 (0.05)**	

3.1. Cross-validation evaluation results

Table 3 summarizes the CT-synthesis quantitative comparison experiment obtained with different CT-synthesis imaging evaluation metrics. We can see from Table 3 that the mean MAE obtained (26.44(0.62) HU) by the proposed feature conditional GAN from the 36 patients in the DS1 dataset using the 6-fold cross-validation strategy is lower than those obtained by the conditional GAN and the Cycle GAN models. The FID, SWD and BD score presented in Table 4 demonstrate the quantitative performance of the three models at the task of generating a variety of sCT images. The obtained metrics were 134.45(0.21) for FID, 7.38(0.01) for SWD, and 0.94(0.05) for BD. Table 5 presents the resulting Dice scores on the bone, the soft tissue and the air areas. Results show that our model produces the highest Dice score compared to the other models in the three segmented sCT images. The bone, soft tissue and the air dice scores correspond respectively to 0.74(0.05), 0.76(0.05) and 0.82 (0.02). The visual comparison presented in figure 3 also shows that the proposed model yielded perceptually different CT-synthesis images that match accurately the different CT-ground truth, contrary to the other reference deep models which are visually less reliable. We can also note that the Conditional GAN model produces qualitatively better synthetic images result than the Cycle GAN. We should also mention that the standard deviations obtained in our study are smaller than those reported in the state-of-the-art methods (Brou Boni *et al* 2020, Maspero *et al* 2020, Peng *et al* 2020), possibly suggesting that a larger dataset is required to estimate statistically a reliable standard deviation which expresses statistically significant difference with other models (Binu *et al* 2014, Ott and Longnecker 2015, Mishra *et al* 2019). Furthermore, the homogeneity of the datasets can also affect the range in standard deviation.

We finally observe from the Tables 6 and 7 that the different models used in this study are able to generate the MRI-T1 images from the CT, in which the performance of our proposed model outperforms the cycle and the conditional GANs with a MAE of 12.28(0.14) and a BD of 0.89(0.02). Through the comparison MRI-synthesis

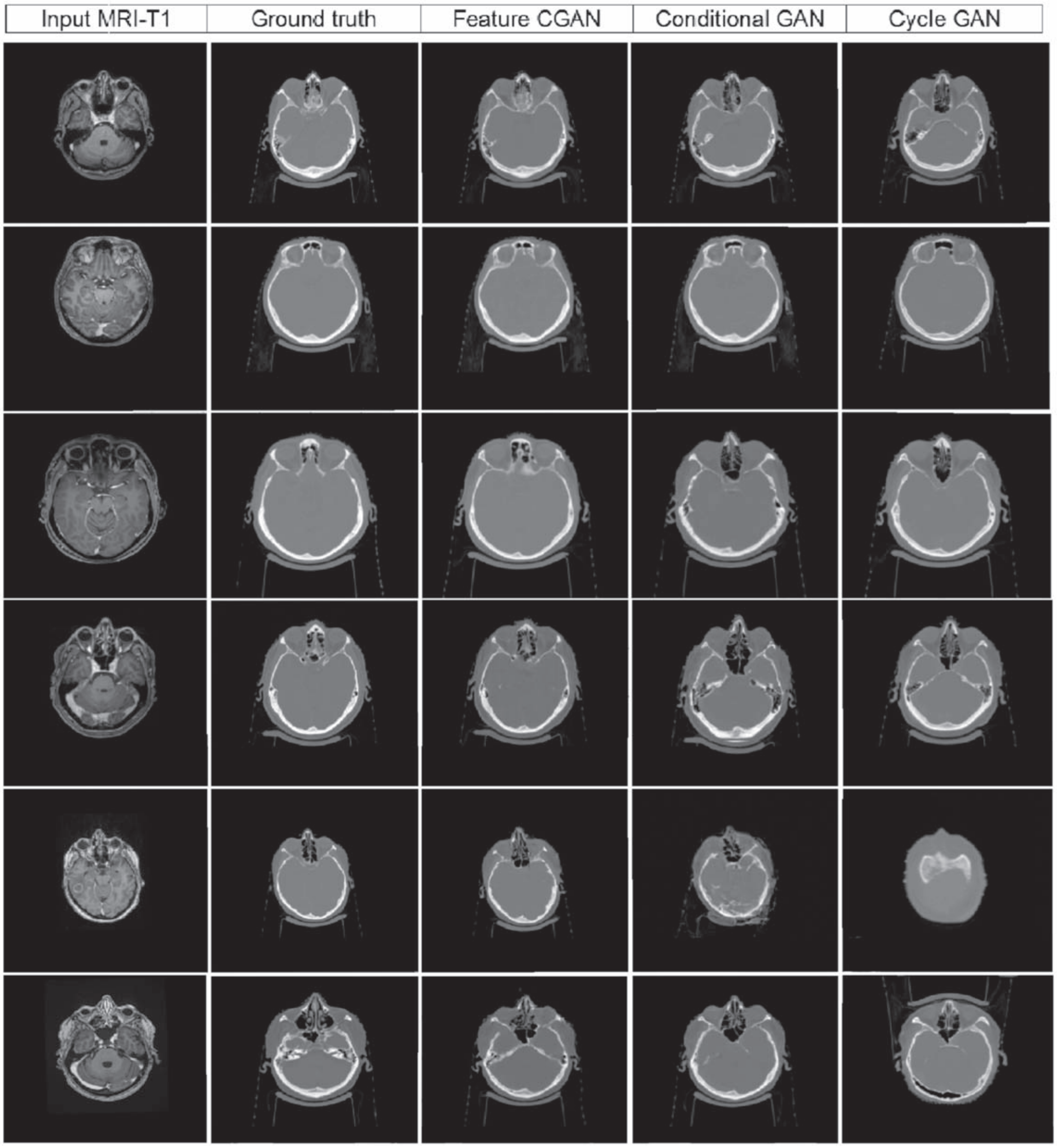


Figure 3. CT-synthesis comparison in the HU space from the DS1 dataset. From left to right: input MRI-T1 slice, ground-truth CT, resulting sCT image of the proposed model, conditional GAN, and CycleGAN.

Table 6. MAE, PSNR, SSIM, and PCC evaluations of MRI-T1 synthesis of the proposed model and the state-of-the-art deep synthesis models, using the head and neck DS1 imaging dataset. (⇓) Lower values indicate improved image synthesis quality. (⇑) Higher values indicate improved image synthesis quality.

	MAE (std) (⇓)	**PSNR (std) DB** (⇑)	**MSSIM index (std)** (⇑)	**PCC (std)** (⇑)
Cycle GAN	25.24 (0.14)	19.08 (1.55)	0.71 (0.05)	0.80 (0.03)
Conditional GAN	18.75 (1.08)	28.31 (1.12)	0.82 (0.07)	0.86 (0.04)
Proposed model	12.28 (0.14)	31.89 (0.12)	0.84 (0.01)	0.87 (0.01)

Table 7. FID, SWD, and BD evaluations of MRI-T1 synthesis of the proposed model and the state-of-the-art deep synthesis models, using the head and neck cross-validation DS1 imaging dataset. (⇓) Lower values indicate improved image synthesis quality. (⇑) Higher values indicate improved image synthesis quality.

	FID (std) (⇓)	**SWD (std)** (⇓)	**BD (std)** (⇑)
Cycle GAN	152.46 (0.38)	11.75 (0.02)	0.85 (0.02)
Conditional GAN	121.74 (0.72)	7.90 (0.04)	0.87 (0.05)
Proposed model	87.29 (0.28)	2.53 (0.07)	0.89 (0.02)

Table 8. MAE, PSNR, SSIM, and PCC evaluations of CT-synthesis of the proposed model and the state-of-the-art deep synthesis models, using the holdout DS2 dataset. (⇓) Lower values indicate improved image synthesis quality. (⇑) Higher values indicate improved image synthesis quality.

	MAE (std) HU (⇓)	**PSNR (std) DB (⇑)**	**MSSIM index (std) (⇑)**	**PCC (std) (⇑)**
Cycle GAN	113.87 (1.21)	25.65 (0.09)	0.79 (0.05)	0.86 (0.01)
Conditional GAN	61.53 (1.15)	32.01 (0.01)	0.88 (0.04)	0.91 (0.07)
Proposed model	45.30 (1.87)	36.37 (0.09)	0.92 (0.05)	0.95 (0.04)

Table 9. FID, SWD, and BD evaluations of CT-synthesis of the proposed model and the state-of-the-art deep synthesis models, using the holdout DS2 dataset. (⇓) Lower values indicate improved image synthesis quality. (⇑) Higher values indicate improved image synthesis quality.

	FID	**SWD**	**BD**
	(std) (⇓)	**(std)** (⇓)	**(std)** (⇑)
Cycle GAN	312 (8.641)	19.23 (1.05)	0.81 (0.06)
Conditional GAN	201 (5.01)	14.42 (1.21)	0.84 (0.07)
Proposed model	176 (7.53)	9.17 (2.84)	0.92 (0.02)

Table 10. Dice score of CT-synthesis on the DS2 imaging dataset obtained by our model and the state-of-the-art synthesis models, through the three segmented areas. (⇑) higher is the value better is the quality synthesis.

	Dice score (std) (⇑)	**Bone area HU**
Cycle GAN	0.56 (0.02)	
Conditional GAN	0.60 (0.07)	(>300)
Proposed model	0.71 (0.03)	
	Dice score (std) (⇑)	**Air area HU**
Cycle GAN	0.76 (0.08)	
Conditional GAN	0.79 (0.01)	(<-100)
Proposed model	0.79 (0.05)	
	Dice score (std) (⇑)	**Soft-tissue area HU**
Cycle GAN	0.65 (0.05)	
Conditional GAN	0.68 (0.04)	$(-100 >$ and $< 300)$
Proposed model	0.75 (0.07)	

results shown in Figure 5, it can be seen that our proposed MRI-synthesis model is more effective than the cycle and conditional GANs synthesis models, for CT to MRI-T1 image synthesis task as well. This can be explained by the fact that our model uses relatively invariant imaging modality representation based on high frequency patterns.

3.2. Independent evaluation results

The final set of experiments was performed on the 20 unseen head and neck images from DS2, used for evaluating the performance of the three synthesis models to synthesize the CT image from the MRI-T1, and inversely to ensure consistency. The obtained CT-synthesis results shown in Tables 8, 9 and 10 demonstrate the learning performances of the different deep models to synthesize a new unseen head and neck image patient from the MRI-T1 images. Based on the different evaluation metrics results, we can observe that the conditional and the Cycle GAN provide lower image quality, with lower scores compared to the proposed model (see Figure 4), in which we obtain the highest Dice score for synthesizing the bone with a score of 0.71(0.03) meanwhile a slightly higher score for the air and soft-tissue areas equal to 0.79(0.05) and 0.75(0.07).

In the inverse processes, Tables 11 and 12 show the learning performances of the presented models to process a new unseen CT image to generate an MRI-T1 image, in which we can notice that the average synthetic results of our approach are better than the other methods with a MAE of 32.0(1.11).

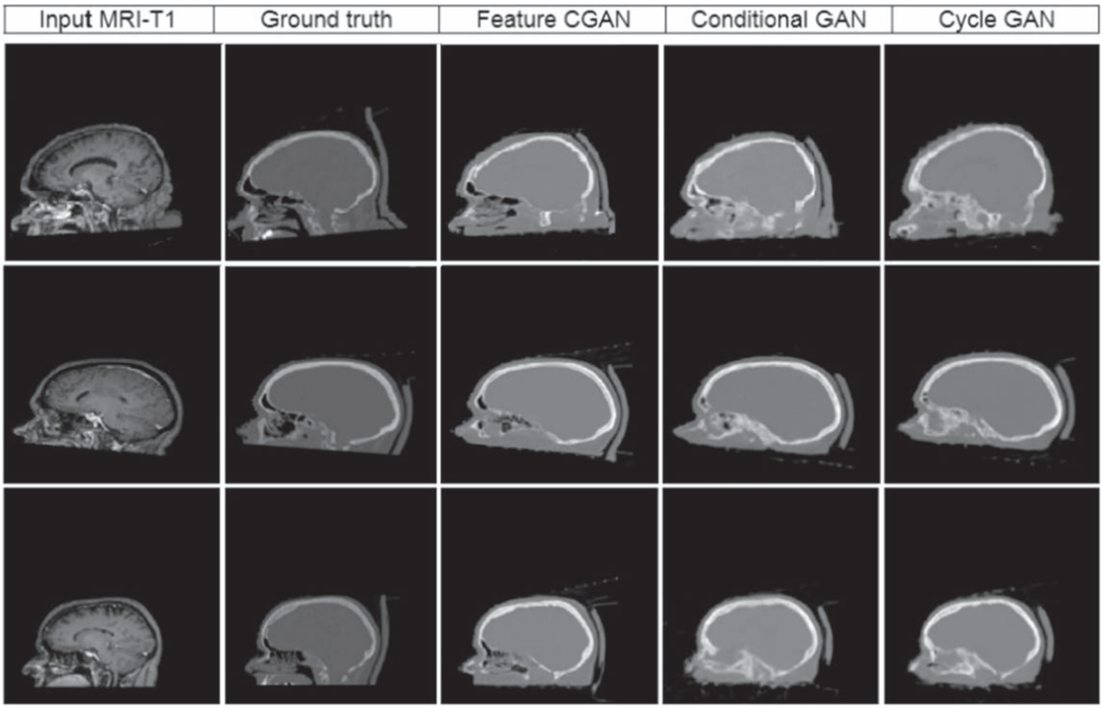


Figure 4. CT-synthesis comparison on the DS2 dataset. From left to right: input MRI-T1 slice, ground-truth, resulting sCT image from the proposed model, conditional GAN, and CycleGAN.

Table 11. MAE, PSNR, SSIM, and PCC evaluations of MRI-T1 synthesis of the proposed model and the state-of-the-art deep synthesis models, using the holdout DS2 imaging dataset. (⇓) Lower values indicate improved image synthesis quality. (⇑) Higher values indicate improved image synthesis quality.

	MAE (std) (⇓)	**PSNR (std) DB (⇑)**	**MSSIM index (std) (⇑)**	**PCC (std) (⇑)**
Cycle GAN	75.84 (1.68)	17.85 (0.04)	0.69 (0.03)	0.79 (0.09)
Conditional GAN	56.52 (1.11)	26.24 (0.09)	0.81 (0.04)	0.84 (0.01)
Proposed model	32.04 (1.11)	29.84 (0.06)	0.83 (0.02)	0.85 (0.05)

Table 12. FID, SWD, and BD evaluations of MRI-T1 synthesis of the proposed model compared with state-of-the-art deep synthesis models, using the holdout DS2 imaging dataset. (⇓) Lower values indicate improved image synthesis quality. (⇑) Higher values indicate improved image synthesis quality.

	FID (std) (⇓)	**SWD (std) (⇓)**	**BD (std) (⇑)**
Cycle GAN	223.43 (0.15)	13.24 (1.25)	0.80 (0.03)
Conditional GAN	167.01 (0.25)	11.05 (1.12)	0.84 (0.02)
Proposed model	103.12 (0.15)	3.82 (1.102)	0.87 (0.01)

4. Discussion

The initial rationale for a feature based synthesis model was to propose a generative framework which could accommodate feature learning in both edge and intensity regions, as opposed to other models which would focus on the predominant features for training. The main difference lies in the feature representation used to guide the generator to produce accurate outputs for CT-synthesis images. The cycle and conditional GAN models use solely and directly the deep feature representation coming from the CNN discriminator. This is opposed to the proposed model which uses a learning representation that integrates complementary information. The proposed representation encodes contour characteristics of the deep feature space resulting from the discrimination phase. Using our representation, the model can achieve improved results that can be observed in both qualitative and quantitative evaluations. The conducted experiments demonstrate that the proposed model outperforms the conditional GAN and the Cycle GAN models under a variety of anatomies and tumor locations.

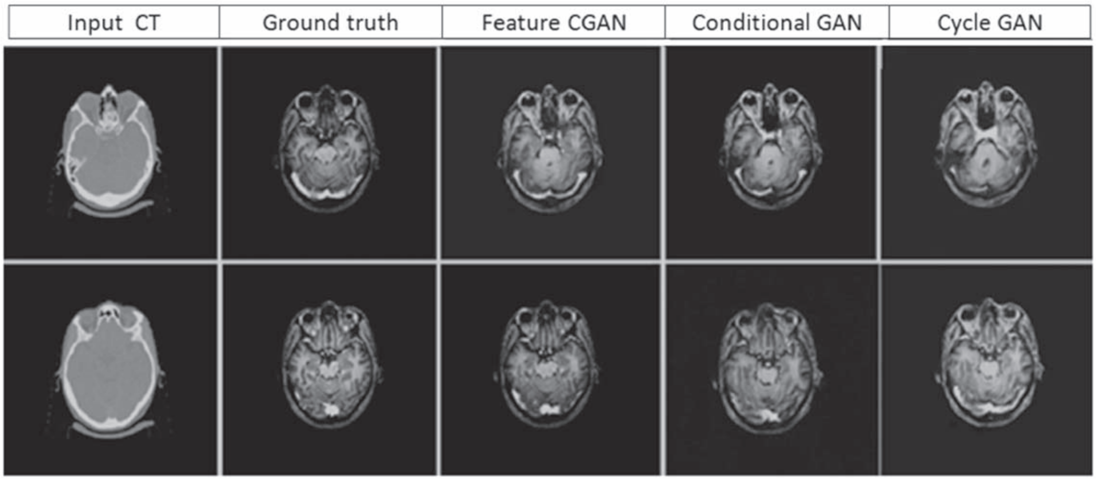


Figure 5. Example of MRI-T1 synthesis result from CT. From left to right: input CT slice, ground-truth MRI-T1, resulting synthetic MRI-T1 image of the proposed model, 2D conditional GAN, and CycleGAN.

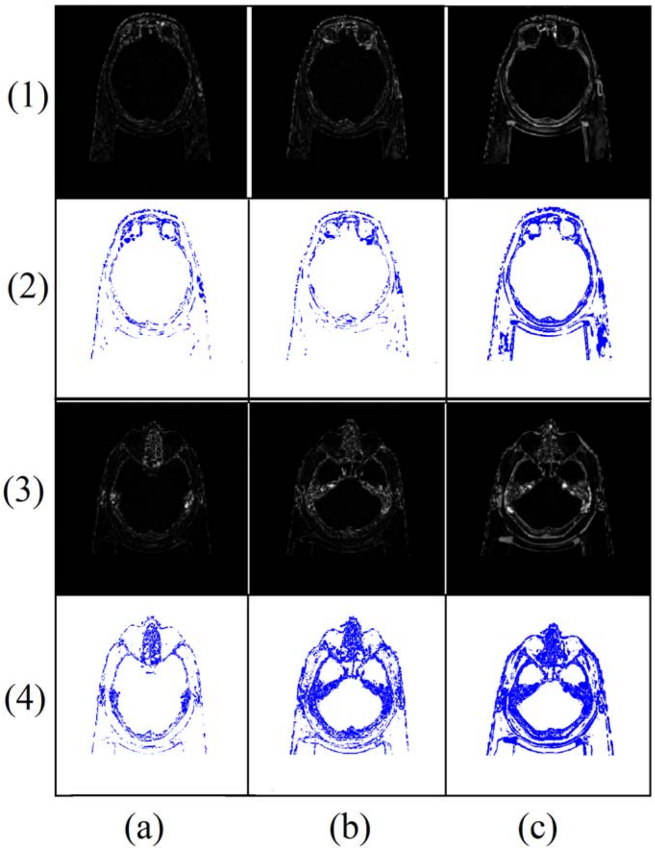


Figure 6. Absolute error maps in the (1), (3) HU space and (2), (4) its corresponding confusion maps. The comparison is made between the (a) proposed model, (b) conditional GAN, (c) Cycle GAN. White and blue colors represent, respectively, the true positives (TP), false positives (FP) and false negatives (FN).

While in some cases, the results from the proposed model are similar to the conditional GAN results, this can be explained by the complexity of the anatomy observed in an image. For example, in the case of the second patient shown in the second row of Figure 3, the presented slice contains normal healthy tissue devoid of tumors. In such cases, the CT-image synthesis can be generated by the three methods with high efficiency. The similarity of the results is then explained by the simplicity of the shapes in such slices, which in some cases fails to exploit the benefits presented by our model. Figure 6 shows the error maps obtained by comparing the CT-synthesis results to the ground truth CT images with accurate HU and their corresponding confusion map obtained by

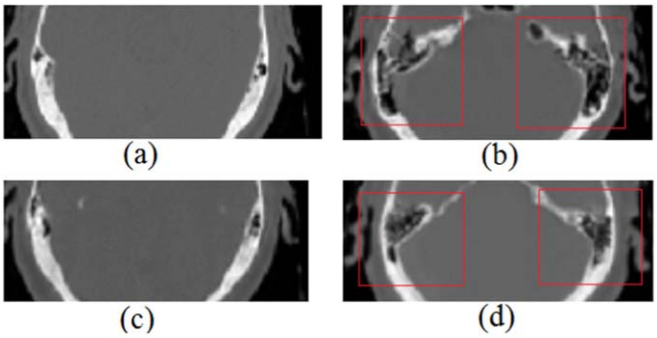


Figure 7. Examples of local image synthesis results, comparing the (a) ground-truth CT, (b) 2D conditional GAN, (c) proposed model, (d) 3D pseudo Cycle GAN. Results shown traditional models failing in skull region, whereas the purposed models captures the transition from skull to the brain region.

encoding the error maps with a binary class labels reflecting the true positives (TP), false positives (FP) and false negatives (FN) signals. Here we encode the false incorrect predictions (FP and FN) with the blue color and TP with white color. The first and the second rows of Figure 6 show a case where the three obtained error maps are similar except in the border of the synthetic images between the objects of the sCT image, where the conditional and the cycle GAN either failed to synthesize the appropriate structures or merged multiple regions in the same structure. The proposed model seems to avoid this behavior and maintains fine details in the shown regions. This means that the feature invariant model succeeds in synthesizing the edges between anatomical structures, with more reliable complete structures due to the histogram contour based feature comparison that indicates the existence of contours surrounding the anatomy.

Furthermore, without loss of generality, the results obtained with the proposed model are comparable to the Cycle GAN and the conditional GAN results for simpler cases, which show less variations in the anatomy. For more complex cases, such as illustrated in Figure 6 (rows 3 and 4), our model provides significant improvements which highlights the difference between our obtained confusion map to those generated by the conditional GAN or the Cycle GAN. The confusion maps generated from the feature conditional GAN expresses a more accurate image synthesis, showing the smallest region of false synthesis. This indicates that the efficiency of the proposed method can be appreciated within more complex regions, where multiple anatomical structures are present, such as is the case of patient 4 (see Figure 3). In these cases, we can observe the existence of false internal regions on the two sides of the object, i.e. false anatomy is mistakenly synthesized (see Figure 7) in the 3D pseudo cycle and in the 2D conditional GAN models. However, our feature conditional GAN results in more accurate outputs and generates lower confusion error maps, by increasing the true positives and reducing the false negative synthesis. This can be interpreted by the fact that our model can learn a more effective representation from the combination of complementary histogram representations, contour and appearance features based on the convolutional map described in the CNN feature space. These features allow us to guide the model to learn the absence as well as the existence of contours (equivalent to the existence or the absence of regions), based on two matching operations of different and complementary feature histograms between the generated CT and the real CT images, capturing the variability of the different anatomies. This considerably increases the performance of the model, as shown again by the improved Dice scores computed on the segmented CT synthesis images, with the segmentation distinguishing three components: bone, soft tissue, and air. The advantage of the proposed method compared to the similar state-of-the-art methods based on the reported Dice values of the three components indicates that the overlapping geometric shape produced by the model offers improved mapping better than the others. In the case of Figures 3 and 4, an artifact can be observed corresponding to a portion of the table which was too close to the skull to be removed by the preprocessing steps. As our training is achieved on the preprocessed data in which this noise represents a very small proportion of the entire image, the training process is thus not affected significantly. Furthermore, these artifacts can be delt with to not affect the subsequent dose calculation steps via postprocessing of the sCT by applying the masks on the region of interest, which are a necessary component of the dose optimization step. The evaluation and the comparison studies show that the synthesis of our conditional GAN model based invariant representation performs better than the conditional GAN and Cycle GAN, which suggests that the choice of the representation learning and the deep neural network architecture both play a crucial role in performance enhancement of the image synthesis.

This study constitutes a preliminary attempt of using clinical MRI examinations with planning CT scans obtained prior to radiotherapy to perform an evaluation based on image quality. In fact, given the target goal of this project to provide an MRI-only workflow for dose plan calculations, dosimetric evaluation of our proposed

CT synthesis method remains to be performed. For this, the construction of a larger benchmark head and neck paired MR/CT dataset with corresponding dose plan parameters is currently being explored. Future work will explore evaluating our proposed approach directly on the downstream task of generating clinically accurate dose plans using sCT images generated from clinical MR images.

5. Conclusion

In this work, we have presented a new and robust CT/MRI synthesis model for radiotherapy planning from head and neck acquisitions. Our model proposes a compact multiscale-invariant representation that encodes the feature space of the conditional GAN discriminator, with complementary features appearing at different image resolution levels. This new invariant representation supports the generator to produce a synthetic image that preserves the specific common appearance and high frequency of CT-image patterns. Qualitative and quantitative results show that our learning strategy is reliable and efficient for generating an accurate sCT image from a set of MR images showing various subject shape sizes with different spatial resolution images. Future work will include training on larger and multicentric datasets, as well as perform a dosimetric evaluation of radiotherapy plans generated from the synthesized images.

Acknowledgments

We would like to thank our funding agencies (NSERC (RGPIN-2020-06558), MEDTEQ, FRQS (293740)) for supporting this research, and the radiation oncology department at the CHUM hospital for the dataset preparation.

Compliance with Ethical Standards

All experiments performed in the study involving human participants were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. Approval was granted by the local Institutional Review Board.

ORCID iDs

Redha Touati https://orcid.org/0000-0003-3845-5361
Samuel Kadoury https://orcid.org/0000-0002-3048-4291

References

Bae S and Kim M 2015 A novel ssim index for image quality assessment using a new luminance adaptation effect model in pixel intensity domain *2015 Visual Communications and Image Processing (VCIP)* pp 1–4

Binu V S, Mayya S S and Dhar M 2014 Some basic aspects of statistical methods and sample size determination in health science research *Ayu* **35** 119

Brou Boni K N D, Klein J, Vanquin L, Wagner A, Lacornerie T, Pasquier D and Reynaert N 2020 Mr to ct synthesis with multicenter data in the pelvic area using a conditional generative adversarial network *Phys. Med. Biol.* **65** 075002

Crum W R, Camara O and Hill D L G 2006 Generalized overlap measures for evaluation and validation in medical image analysis *IEEE Trans. Med. Imaging* **25** 1451–61

Dalal N and Triggs B 2005 Histograms of oriented gradients for human detection *IEEE comp. soc. conf. on computer vis. and patt. recog. (CVPR'05)* **vol 1** 886–93

Deshpande I, Zhang Z and Schwing A G 2018 Generative modeling using the sliced wasserstein distance *IEEE/CVF Conf. on Comp. Vision and Patt. Recog. (CVPR* **2018** 3483–91

Dong N, Roger T, Jun L, Caroline P, Su R, Qian W, Dinggang S and Simon D 2017 Medical image synthesis with context-aware generative adversarial networks *Med. Image Comp. and Comput. Assist. Intervention, MICCAI 2017 (20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part III* 10435) ed M. Descoteaux, L Maier-Hein, A Franz, P Jannin, D L Collins and S Duchesne 1 edn (Berlin: Springer) pp 417–25

Dowling J A, Lambert J, Parker J, Salvado O, Fripp J, Capp A, Wratten C, Denham J W and Greer P B 2012 An atlas-based electron density mapping method for magnetic resonance imaging (mri)-alone treatment planning and adaptive mri-based prostate radiation therapy *Int. J. Radiat. Oncol. Biol. Phys.* **83** e5–11

Frangi A F, Tsaftaris S A and Prince J L 2018 Simulation and synthesis in medical imaging *IEEE Trans. Med. Imaging* **37** 673–9

Gonzalez R C and Woods R E 2018 *Digital Image Processing* 4th edn (New York: Pearson)

Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A and Bengio Y 2014 Generative adversarial nets *NIPS'14: Proc. of the 27th Int. Conf. on Neur. Info. Proces. Systems* (Canada: MIT Press) pp 2672–80

Hattangadi J A, Chapman P, Kim D, Bussiere M, Niemierko A, Rowell A, Daartz J, Ogilvy C, Loeer J and Shih H 2012 single fraction proton beam stereotactic radiosurgery (psrs) for inoperable cerebral arteriovenous malformations (avms) *Int. J. Radiat. Onocol. Biol. Phys.* **84** S38

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 2016 Deep residual learning for image recognition *2016 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* pp 770–8

Heusel M, Ramsauer H, Unterthiner T, Nessler B and Hochreiter S 2017 Gans trained by a two time-scale update rule converge to a local nash equilibrium *NIPS'17: Proc. of the 31st Int. Conf. on Neural Info. Proc. Syst.* (Long Beach: Curran Associates) pp 6626–37

Hofmann M, Steinke F, Scheel V, Charpiat G, Farquhar J, Aschoff P, Brady M, Schölkopf B and Pichler B J 2008 Mri-based attenuation correction for pet/mri: a novel approach combining pattern recognition and atlas registration *J. Nucl. Med.* **49** 1875–83

Huynh T, Gao Y, Kang J, Wang L, Zhang P, Lian J and Shen D 2015 Estimating ct image from mri data using structured random forest and auto-context model *IEEE Trans. Med. Imaging* **35** 174–83

Huynh T, Gao Y, Kang J, Wang L, Zhang P, Lian J and Shen D 2016 Estimating ct image from mri data using structured random forest and auto-context model *IEEE Trans. Med. Imaging* **35** 174–83

Işın A, Direkoğlu C and Şah M 2016 Review of mri-based brain tumor image segmentation using deep learning methods *Procedia Comput. Sci.* **102** 317–24

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

Johansson A, Garpebring A, Karlsson M, Asklund T and Nyholm T 2013 Improved quality of computed tomography substitute derived from magnetic resonance (mr) data by incorporation of spatial information-potential application for mr-only radiotherapy and attenuation correction in positron emission tomography *Acta Oncol.* **52** 1369–73

Jonsson J H, Johansson A, Söderström K, Asklund T and Nyholm T 2013 Treatment planning of intracranial targets on mri derived substitute ct data *Radiother. Oncol.* **108** 118–22

Kailath T 1967 The divergence and bhattacharyya distance measures in signal selection *IEEE Trans. Commun. Technol.* **15** 52–60

Kazemifar S, Barragán Montero A M, Souris K, Rivas S T, Timmerman R, Park Y K, Jiang S, Geets X, Sterpin E and Owrangi A 2020 Dosimetric evaluation of synthetic ct generated with gans for mri-only proton therapy treatment planning of brain tumors *J. appl. clin. med. phys.* (https://doi.org/10.1002/acm2.12856)

Keereman V, Fierens Y, Broux T, De Deene Y, Lonneux M and Vandenberghe S 2010 Mri-based attenuation correction for pet/mri using ultrashort echo time sequences *J. Nucl. Med.* **51** 812–8

Kingma D P and Ba J 2015 Adam: A method for stochastic optimization, 2014 *The 3rd Int. Conf. for Learning Representations* (San Diego)

Klages P, Benslimane I, Riyahi S, Jiang J, Hunt M, Deasy J O, Veeraraghavan H and Tyagi N 2020 Patch-based generative adversarial neural network models for head and neck mr-only planning *Med. Phys.* **47** 626–42

Korsholm M E, Waring L W and Edmund J M 2014 A criterion for the reliable use of mri-only radiotherapy *Radiat. Oncol.* **9** 16

Lauritzen A D, Papademetris X, Turovets S and Onofrey J A 2019 Evaluation of ct image synthesis methods:from atlas-based registration to deep learning arXiv:1906.04467

Li Y, Li W, He P, Xiong J, Xia J and Xie Y 2019 Ct synthesis from mri images based on deep learning methods for mri-only radiotherapy *2019 Int. Conf. on Medical Imaging Physics and Engineering (ICMIPE)* **2019** pp 1–6

Li Y, Zhu J, Liu Z, Teng J, Xie Q, Zhang L, Liu X, Shi J and Chen L 2019 A preliminary study of using a deep convolution neural network to generate synthesized CT images based on CBCT for adaptive radiotherapy of nasopharyngeal carcinoma *Phys. Med. Biol.* **64** 145010

Liang X, Chen L, Nguyen D, Zhou Z, Gu X, Yang M, Wang J and Jiang S 2019 Generating synthesized computed tomography (CT) from cone-beam computed tomography (CBCT) using CycleGAN for adaptive radiation therapy *Phys. Med. Biol.* **64** 125002

Litjens G, Kooi T, Bejnordi B E, Setio A A A, Ciompi F, Ghafoorian M, Van Der Laak J A, Van Ginneken B and Sánchez C I 2017 A survey on deep learning in medical image analysis *Med. Image Anal.* **42** 60–88

Liu L, Johansson A, Cao Y, Dow J, Lawrence T S and Balter J M 2020 Abdominal synthetic ct generation from mr dixon images using a u-net trained with emi-synthetic'ct data *Phys. Med. Biol.* **65** 125001

Liu Y *et al* 2019 MRI-based treatment planning for proton radiotherapy: dosimetric validation of a deep learning-based liver synthetic CT generation method *Phys. Med. Biol.* **64** 145015

Maspero M, Bentvelzen L G, Savenije M H F, Guerreiro F, Seravalli E, Janssens G O, van den Berg C A T and Philippens M E P 2020 Deep learning-based synthetic ct generation for paediatric brain mr-only photon and proton radiotherapy *Radiother. Oncol.* 197–204

Metcalfe P, Liney G P, Holloway L, Walker A, Barton M, Delaney G P, Vinod S and Tome W 2013 The potential for an enhanced role for mri in radiation-therapy treatment planning *Technol. Cancer Res. Treat.* **12** 429–46

Milletari F, Navab N and Ahmadi S-A 2016 V-net: Fully convolutional neural networks for volumetric medical image segmentation *2016 4th Int. conf. on 3D Vision (3DV)* pp 565–71

Mishra P, Pandey C M, Singh U, Keshri A and Sabaretnam M 2019 Selection of appropriate statistical methods for data analysis *Ann Card Anaesth* **22** 297–301

Nie D, Cao X, Gao Y, Wang L and Shen D 2016 Estimating ct image from mri data using 3d fully convolutional networks *In Deep Learning and Data Labeling for Medical Applications* (Berlin: Springer) pp 170–8

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 *Int. Conf. on Medical Image Computing and Computer-Assisted Intervention* pp 417–25

Niemierko A, Rowell J, Daartz C, Loeffler O J and Shih H 2012 Single fraction proton beam stereotactic radiosurgery (psrs) for inoperable cerebral arteriovenous malformations (avms) *Int. J. Radiat. Oncol. Biol. Phys.* **84** S38

Ott R L and Longnecker M T 2015 *An Introduction to Statistical Methods and Data Analysis*

Oulbacha R and Kadoury S 2020 MRI to CT synthesis of the lumbar spine from a pseudo-3d cycle GAN *17th IEEE International Symposium on Biomedical Imaging, ISBI 2020, Iowa City, IA, USA, April 3-7, 2020* pp 1784–1787 IEEE

Peng Y *et al* 2020 Magnetic resonance-based synthetic computed tomography images generated using generative adversarial networks for nasopharyngeal carcinoma radiotherapy treatment planning *Radiother. Oncol.* **150** 217–24

Prabhakar R, Julka P K, Ganesh T, Munshi A, Joshi R C and Rath G K 2007 Feasibility of using mri alone for 3d radiation treatment planning in brain tumors *Japan. j. clin. oncol.* **37** 405–11

Qi M *et al* 2020 Multi-sequence mr image-based synthetic ct generation using a generative adversarial network for head and neck mri-only radiotherapy *Med. Phys.* **47** 1880–94

Rank C M, Tremmel C, Hünemohr N, Nagel A M, Jäkel O and Greilich S 2013 Mri-based treatment plan simulation and adaptation for ion radiotherapy using a classification-based approach *Radiat. Oncol.* **8** 51

Robson M D, Gatehouse P D, Bydder M and Bydder G M 2003 Magnetic resonance: an introduction to ultrashort te (ute) imaging *J. Comput. Assist. Tomogr.* **27** 825–46

Ronneberger O, Fischer P and Brox T 2015 U-net: Convolutional networks for biomedical image segmentation *Medical Image Computing and Computer-Assisted Intervention (MICCAI), volume 9351 of LNCS* (Berlin: Springer) pp 234–41 arXiv:1505.04597 [cs.CV]

Stanescu T, Jans H S, Pervez N, Stavrev P and Fallone B G 2008 A study on the magnetic resonance imaging (mri)-based radiation treatment planning of intracranial lesions *Phys. Med. Biol.* **53** 3579

Van der Bom M J, Pluim J P W, Gounis M J, van de Kraats E B, Sprinkhuizen S M, Timmer J, Homan R and Bartels L W 2011 Registration of 2d x-ray images to 3d mri by generating pseudo-ct data *Phys. Med. Biol.* **56** 1031

Wolterink J M, Dinkla A M, Savenije M H F, Seevinck P R, van den Berg C A T and Išgum I 2017 Deep mr to ct synthesis using unpaired data *In International workshop on simulation and synthesis in medical imaging* (Berlin: Springer) pp 14–23

Yi Z, Zhang H, Tan P and Gong M 2017 Dualgan: Unsupervised dual learning for image-to-image translation *Proceedings of the IEEE int. conf. on computer vision* pp 2849–57

Zhong G, Wang L-N, Ling X and Dong J 2016 An overview on data representation learning: From traditional feature learning to recent deep learning *J. Finance Data Sci.* **2** 265–78

Zhu J-Y, Park T, Isola P and Efros A A 2017 Unpaired image-to-image translation using cycle-consistent adversarial networks *Proc. of the IEEE int. conf. on computer vision* pp 2223–32