# Multi-planar dual adversarial network based on dynamic 3D features for MRI-CT head and neck image synthesis

IOP Publishing Phys. Med. Biol. 69 (2024) 155012 https://doi.org/10.1088/1361-6560/ad611a

Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

OPEN ACCESS

RECEIVED
23 April 2024

REVISED
18 June 2024

ACCEPTED FOR PUBLICATION
9 July 2024

PUBLISHED
19 July 2024



PAPER

# Multi-planar dual adversarial network based on dynamic 3D features for MRI-CT head and neck 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:** image generation, dynamic features, 3D multi-view image modeling, dual feature learning, adversarial network, generative network model

## Abstract

*Objective.* Head and neck radiotherapy planning requires electron densities from different tissues for dose calculation. Dose calculation from imaging modalities such as MRI remains an unsolved problem since this imaging modality does not provide information about the density of electrons. *Approach.* We propose a generative adversarial network (GAN) approach that synthesizes CT (sCT) images from T1-weighted MRI acquisitions in head and neck cancer patients. Our contribution is to exploit new features that are relevant for improving multimodal image synthesis, and thus improving the quality of the generated CT images. More precisely, we propose a Dual branch generator based on the U-Net architecture and on an augmented multi-planar branch. The augmented branch learns specific 3D dynamic features, which describe the dynamic image shape variations and are extracted from different view-points of the volumetric input MRI. The architecture of the proposed model relies on an end-to-end convolutional U-Net embedding network. *Results.* The proposed model achieves a mean absolute error (MAE) of $18.76 \pm 5.167$ in the target Hounsfield unit (HU) space on sagittal head and neck patients, with a mean structural similarity (MSSIM) of $0.95 \pm 0.09$ and a Frechet inception distance (FID) of $145.60 \pm 8.38$. The model yields a MAE of $26.83 \pm 8.27$ to generate specific primary tumor regions on axial patient acquisitions, with a Dice score of $0.73 \pm 0.06$ and a FID distance equal to $122.58 \pm 7.55$. The improvement of our model over other state-of-the-art GAN approaches is of 3.8%, on a tumor test set. On both sagittal and axial acquisitions, the model yields the best peak signal-to-noise ratio of $27.89 \pm 2.22$ and $26.08 \pm 2.95$ to synthesize MRI from CT input. *Significance.* The proposed model synthesizes both sagittal and axial CT tumor images, used for radiotherapy treatment planning in head and neck cancer cases. The performance analysis across different imaging metrics and under different evaluation strategies demonstrates the effectiveness of our dual CT synthesis model to produce high quality sCT images compared to other state-of-the-art approaches. Our model could improve clinical tumor analysis, in which a further clinical validation remains to be explored.

## 1. Introduction

Medical imaging synthesis is an automatic process that can be used for transforming one imaging modality to a target imaging modality, which faithfully represents the same anatomical structures as the source image (Frangi *et al* 2018, Abu-Srhan *et al* 2021). Because patients can sometimes be scanned with systems with different physical properties, the synthesis process should generate a mapping under the assumption that while the original and target images represent different statistical image space with highly heterogeneous appearance characteristics, any existing particularities such as lesions and tumors ought to be preserved (Li *et al* 2019, Liu 2019, Touati *et al* 2023). This challenging task is often a crucial step for different clinical applications, including radiotherapy treatment planning, tumor volume localization, clinical pathology assessment or image registration (Purdy *et al* 2012). These applications often benefit from the

complementary multi-modal information provided by magnetic resonance imaging (MRI) and computed tomography (CT) images to improve pathology diagnosis and consequently disease treatment (Ninon Burgos *et al* 2015, Li *et al* 2019, Liang *et al* 2019, Liu 2019, Kazemifar *et al* 2020, Oulbacha and Kadoury 2020, Abu-Srhan *et al* 2021, Touati *et al* 2021). In radiotherapy, the treatment is dependent on the location and the stage of the cancer: for example, given an imaging modality such as MRI that shows tumors at early stages, the therapeutic objective consists in decreasing toxicity through restricting the dose and the target volume. CT images are also used for performing dose delivery calculation as they provide the required electron density information of the tissues in Hounsfield units (HU) (Purdy *et al* 2012). Meanwhile MRI provides a higher tissue contrast as well as being well adapted for identifying soft tissue tumors (Purdy *et al* 2012). Clinical improvements in CT-dependent dose planning could be achieved with a method exploiting the improved soft-tissue contrast MRI provides, leading to more accurate delineations of OARs, and thus improve the overall quality of dose delivery. Furthermore, such a solution using image synthesis provides additional benefits to the patient: the removal of ionization exposure with CT scans and the lower costs from only acquiring a single MRI scan.

## 2. Related works

Many approaches have been proposed to address the MRI to CT translation problem in recent years. These can be classified into three groups: density techniques, single or multiple atlas methods, and machine learning approaches.

**1) Bulk density assignment techniques.** These methods uses morphological operations with a prior threshold to delineate the volume of interest following the label class of the tissue. After the thresholding stage, a physical or electron density value for each region would be assigned, defining the final synthetic sCT image (Keereman *et al* 2010, Hattangadi 2012, Rank *et al* 2013). These methods suffer mainly from the need of manual intervention to tune the appropriate intensity threshold and to select the suitable range of density values, often in an anatomical region or for specific applications.

**2) Atlas based methods.** The methods based on atlases focus mainly on registration techniques between the MRI and CT scans, using structural similarity measures between the generated atlas and the target patient (Stanescu *et al* 2008, Johansson *et al* 2013). Multi-atlas techniques fuse several atlases after a registration process with the MRI scan (Prabhakar *et al* 2007, Jonsson *et al* 2013, Korsholm *et al* 2014). However, registration errors are highly sensitive to variation between patients due to atypical anatomies, to the optimal atlas count used in the registration stage, as well as the choice of a suitable image similarity metric (Kazemifar *et al* 2020).

**3) Machine learning based methods.** Machine learning models are approaches that use a learning criterion to transform voxels from the intensity space of the MRI to the CT image space, using a generative model for density assignment combined with a classification framework for synthetic image evaluation (Robson *et al* 2003, Van der Bom *et al* 2011, Dowling *et al* 2012, Hattangadi 2012, Metcalfe *et al* 2013, Huynh *et al* 2015). Deep learning approaches particularly, which is a subset of machine learning, take advantage of large computational power to have the model learn the suitable internal representation of the anatomy for the task at hand, while taking into account variations that naturally occur in the large sample population (Goodfellow *et al* 2016).

In recent years, research in medical image diagnosis (Işın *et al* 2016, Litjens *et al* 2017), detection applications (Abbasian Ardakani *et al* 2020, Ozturk *et al* 2020, Narin 2021), or directly in image synthesis itself have demonstrated that high fidelity sCTs are achievable with deep learning methods (Frangi *et al* 2018, Abu-Srhan *et al* 2021). Generative adversarial networks (GANs) in particular are a class of neural network that are also often used in domain translation tasks (Goodfellow *et al* 2014). They also shown improved performance when using paired imaging modalities (Nie *et al* 2017, Wolterink *et al* 2017, Abu-Srhan *et al* 2021) which occurs when patients get multiple scans done for different clinical objectives. The U-Net (Ronneberger *et al* 2015) based GAN was used to improve head and neck CT synthesis from MRI (Dinkla *et al* 2019, Klages *et al* 2020, Qi 2020, Touati *et al* 2021). Improved image synthesis quality was also demonstrated to improve downstream image guidance tasks with the use of 3D Cycle GAN (Zhu *et al* 2017, Oulbacha and Kadoury 2020), DualGAN (Zili *et al* 2017), and 3D fully convolutional network (FCN) (Nie *et al* 2016) for aligned/unaligned MRI images of the spinal (Brou Boni *et al* 2020, Maspero *et al* 2020, Peng 2020), pelvic (Han 2017) and brain area (Dong *et al* 2017). Nevertheless, synthesis of the CT from MRI remains a challenging task in head and neck cases, due to its complex anatomical structures, as well as highly variable and fine-detailed anatomical boundaries (Klages *et al* 2020, Touati *et al* 2021).

Previous works were evaluated on tumor sites such as the brain and prostate (Dong *et al* 2017, Han 2017). Few works were developed for sCT image generation in normal or tumoral head and neck anatomy. Although conditional and cycle GANs (Nie *et al* 2017, Wolterink *et al* 2017, Dinkla *et al* 2019, Klages *et al*

2020, Qi 2020) provide promising results, they often minimize a combination of standard L2 or L1 terms for an adversarial loss, which may also lead to synthesize a blurry image and loss of details. In addition, the different works were designed to synthesize sCT images from a single acquisition plane, without considering augmented similar features (or augmented multiscale features) to generate relatively similar invariant characteristics. To address this issue, a feature-based GAN (Touati *et al* 2021) was proposed to reinforce the generator to generate a sCT image that relatively preserves the boundaries orientation in the generated sCT. However, the feature GAN does not take into account the 3D spatial information when generating the synthetic CT. Ozbey *et al* (2023) proposed a fast adversarial inference diffusion model based on a cycle-consistent learning strategy, using a GAN network in the reverse diffusion step to progressively generate an enhanced denoised image, for cross-modality translation problem with unpaired training data. Dalmaz *et al* (2022) proposed a conditional adversarial training method, using a generator based on a vision transformer and a convolutional branch, for multi-contrast MRI image synthesis and MRI-CT image translation problems. The method employs novel residual transformer blocks in the vision transformer and convolutional branches with a weight sharing strategy, in order to lower the model complexity and preserve local and global information of the generated modality. Yurt *et al* (2021) proposed a multi-level transfer generative framework (msutGAN) for generating missing or corrupted MRI contrasts, using a mixture of one-to-one and many-to-one branches for multi-level features descriptors extraction. These descriptors are then merged in an adversarial training manner to generate the output modality. Askin *et al* (2022) proposed a fast plug-and-play learning approach (PP-MPI) for magnetic particle imaging reconstruction (MPI). The authors pre-trained an image prior for denoising and later embedding it in an alternating direction method of multiplier (ADMM) optimizer for MPI reconstruction, where the trained network iteratively projects the image to regularize the reconstruction. Despite their performances, PP-MPI is affected by scale drifts between training and test images, while msutGAN limits expressiveness for contextual features that reflect contextual dependencies across both healthy and pathological tissues. Hsu *et al* (2022) proposed a multi-planar training method that relies on three orthogonal planes (axial, sagittal and coronal) from paired MR-CT images, for radiotherapy planning in prostate cancer. The method extracts three MRI sets of axial, sagittal and coronal planes from the original acquisitions. The three extracted sets are then used as inputs to the generator for generating axial, sagittal and coronal synthetic sCTs, which are then combined to obtain the final sCT. Moreover, the training loss is an augmented conditional generative adversarial loss, and is composed of an adversarial loss, a pixel reconstruction loss and a mutual information loss. Compared to the multi-planar method of Hsu *et al* (2022), our method proposes a multi-planar representation that captures the dynamic intensity and shape orientation, due in part to the intensity and magnitude gradient features extracted from the axial, sagittal and coronal views. This augmented representation is then fed to the multi-planar encoder branch, while the original image is fed to the encoder U-net stream. The latent features from each stream are combined to generate a head and neck sCT image.

To tackle the issues with CT image synthesis in head and neck cancer cases, we propose in this work a multi-planar dual GAN that estimates sCT images from T1-weighted MRI acquisitions in head and neck cancer cases. Specifically, our approach was designed with the goal to maintain high specificity in preserving cancer lesions, for subsequent therapeutic objectives. The discriminator model is based on the feature invariant GAN (Touati *et al* 2021), for its ability to make use of multiscale high frequency boundary patterns and thus guide the learning process of the generator. Our novel dual branch generator is based on the U-Net with a standard 2D MRI to sCT synthesis branch; it is also augmented with a secondary multi-planar branch that extracts dynamic features along 3 axis of view for the input volumetric MRI in the encoder portion. In the decoder portion of the U-Net, features extracted from both branches are joined to synthesize the final CT image. This approach enhances the generator with 3D information to improve the positioning of anatomical structures. We validate our model on two head and neck cancer datasets with paired CT and T1-weighted MRI obtained from a single institution: one of sagittal acquisitions and one of axial acquisitions. The performance of our dual feature model is then compared to three state-of-the-art image translation methods: CGAN (Nie *et al* 2017), Cycle GAN (Wolterink *et al* 2017) and feature invariant GAN (Touati *et al* 2021). Our work's main novelty is the embedding of 3D dynamic dual features in the GAN model using multi-planar views. These dynamic features allow generating an augmented latent space that captures dynamic structure variations and orientations so that the decoder learns to decode an sCT from an augmented dual space describing local and global image context. Our approach is built on top of the current GAN framework, with experiments showing that the proposed novel dynamic 3D features within the generative model brings a significant improvement in terms of synthesis results and Dice scores. To evaluate our proposed model, a 5-fold cross-validation was performed on the first dataset, and a tumor evaluation was performed on the second dataset, both using a combination of image statistics metrics and Hounsfield tissue thresholding overlap metrics. We also assess the models capability of preserving specific tumor regions during its synthesis process using a Dice overlap score from the expert labeled data.

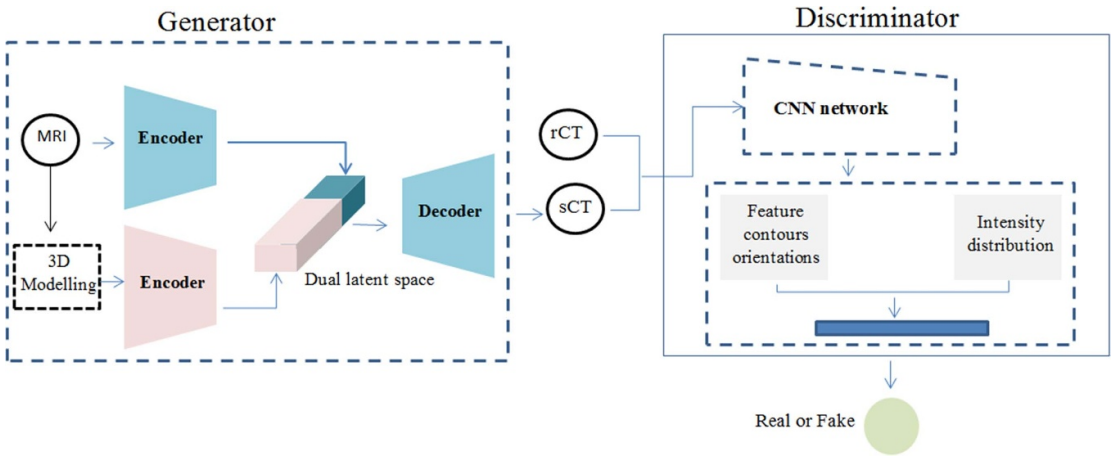


**Figure 1.** General scheme of the proposed dual feature GAN model. In the proposed generator, the convolutional encoder encodes the 2D MRI input (blue color). The embedding encoder maps the 3D representation of the 2D MRI input (pink color), which captures the specific 3D dynamic features. The resulting representation pairs are then concatenated, forming the dual encoding latent space. The decoder (blue color) synthesizes a sCT image from the dual encoding latent space. Once the sCT is generated, the real (rCT) and the sCT images are fed to the discriminator. The CNN network maps the rCT and sCT images via its layers. The resulted convolutional maps are described based on the intensity distribution and the edge feature orientations, then compared to classify them as real or fake.

## 3. Methods

Our proposed network architecture is composed of a dual branch 2D and multi-planar generator network (G) integrating dual feature representation learning, and a discriminator network (D) (see figure 1). For a given pair of co-registered MRI-T1/CT volumes, the multi-planar deep generator (G) ensures the generation of the sCT image from the dual encoding latent feature space considers not only the 2D query image features, but also captures the 3D information by modelling different 2D planar views of the volumetric input data. The discriminator (D) classifies the resulting sCT image into two classes, real or fake, using a multiscale classifier based on previous works using high frequency appearance patterns of the CT image (Touati *et al* 2021).

### 3.1. Multimodal datasets

We train our dual CT-synthesis GAN model on two separate retrospective head-and-neck cancer datasets provided by the radiation oncology department at the Centre Hospitalier de l'Université de Montréal. The first dataset (DS1) is composed of 225 sagittal acquisitions of paired T1-weighted MRI and CT scans. These were obtained between 2015 and 2019 using a 3 T Siemens Magnetom with an image resolution of 0.8mm. The second dataset (DS2) contains 241 axial acquisitions of the same modalities, but with the Philips Achieva 3 T clinical scanner. This resolution for this dataset was mixed, between 0.68 and 0.98 mm. Of the 241 patients, 41 of them contained the primary gross tumor volume (GTV) segmentation, performed by an experienced radiation oncologist.

In both cases, the 3D MRI-T1 images were produced with gradient echo sequences (GRE) and segmented-k-space (SK), while the 3D CT images were acquired with energy levels between 120 and 140 kVp, a repetition time (TR) of 4000 ms, and an echo time (TE) of 12 ms. All MR and CT sequences were rigidly co-registered by a radiation-oncology specialist and were obtained at size $512 \times 512 \times d$ before being cropped down to $256 \times 256 \times 70$ during preprocessing to reduce the amount of noise and blank space present around the edge of the image.

### 3.2. Multi-planar dynamic feature extraction

While a single 2D MR image to be transformed into sCT contains all the necessary intensity information for a synthesis step, it does not take advantage of the spatial context of the slice in 3D space with respect to the rest of the available scanning volume. To encode this, we exploit multiple plane of views (e.g. axial, sagittal and coronal) and model dynamic features for each view: both intensities features as well as image gradient features are computed (see figure 2). Considering all three planes instead of a single one allows information to be processed while maintaining the spatial relationship between each point of view. This preprocessing step allows extracting a set of discriminative image features of the volumetric input data as 2D slices to be exploited during the learning process (see algorithm 1).

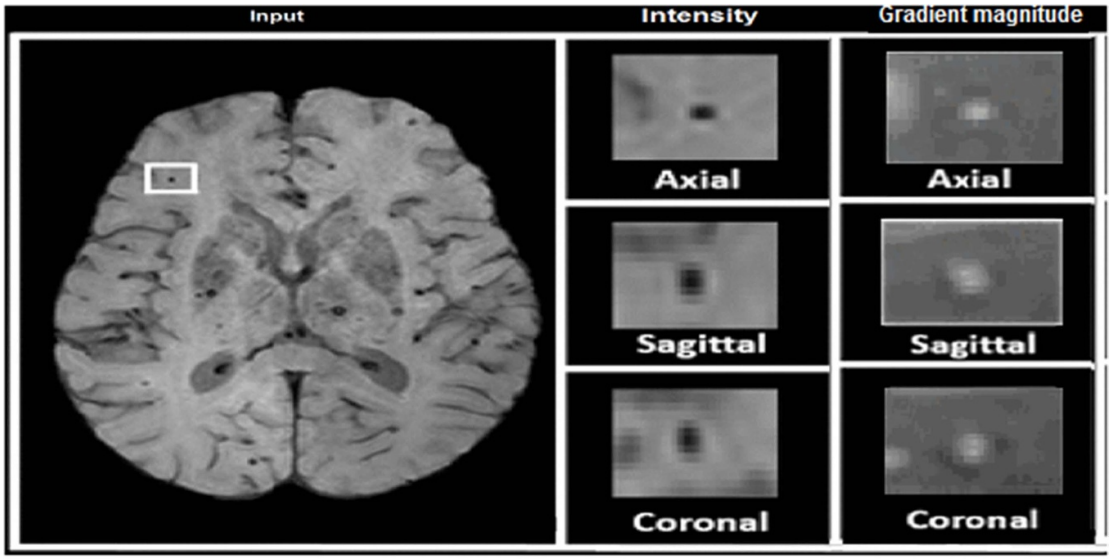


**Figure 2.** Set of features extracted through the axial, sagittal and coronal views. The input is an axial slice of an MRI-T1 image, with a 2D intensity highlighted in white. For each view, the view shows the extracted image pattern and its corresponding magnitude for the 2D intensity in question.

**Algorithm 1.** Multi-planar feature extraction.

Input: $X$ ▷A 3D MRI volume of size $(256, 256, 70)$
Input: $i$ ▷The index of the selected slice for image synthesis
Input: $W$ ▷A neighborhood window of size $(11, 11, 11)$
Output: $F'$ ▷A dynamic feature vector of size $(256, 256, 11 \times 11 \times 11 \times 6)$
**for** $X' \leftarrow X^a; a = axial, sagittal, coronal$ ▷The reoriented 3D MRI volume
  **for** $j, k \leftarrow coordinates(X'_i)$
    $w \leftarrow W(i, j, k)$
    $f[j, k] \leftarrow flatten(X'[w])$ ▷MRI intensity features
    $g[j, k] \leftarrow flatten(\nabla X'[w])$ ▷Gradient magnitude features
  **end**
  $F' \leftarrow F' + f + g$
**end**

First, given a single 3D MRI scan $X$ of dimension $H \times W \times D$, as well as the input 2D MRI $X_i$ to be synthesized (e.g. an axial slice of size $H \times W$), we chose a plane of view $a$ from the available set of axes in 3D space $A = \{axial, sagittal, coronal\}$. Then, for a given pixel position $p$ in the image $X_i$, we consider the neighborhood $W_{h \times w \times d}$ centered around $p$ within the original full 3D scanning volume. This 3D patch, representing the spatial context around the given point $W_p$, is then flattened into a single vector of length $h * w * d$ for every point in the image $X_i$. The resulting feature vector is a 2D image with the channel dimension containing the computed intensity features. The dimension of this output intensity feature ${F^a}_i$ is $H \times W \times (h * w * d)$. This process is repeated for each plane of views $a \in A$: the output feature vector $F$ is the concatenation of these three orientation-specific intensity features. We model the 2D intensity from a 3D cubic window that captures the dynamic variation of this intensity via gradient and intensity information. Each view of the cubic window creates a $W$ series of images ($W$ is the window size)

$$||\nabla F|| = \sqrt{\left(\frac{\delta F}{\delta h}\right)^2 + \left(\frac{\delta F}{\delta w}\right)^2}. \quad (1)$$

This feature extraction strategy is computed for both raw pixel intensities in the input 3D scan, but also for the gradients of the image. For the multi-planar branch, the 2D MRI slice is first transformed in a high multi-dimensional dynamic shape feature space using a multi-planar 3D image representation. An $11 \times 11 \times 11$ patch around each point of the input 2D slice is flattened into an intensity and gradient magnitude feature vector, which is repeated for each of the three possible views (e.g. axial, sagittal and coronal). The resulting multi-planar feature descriptor consists of a feature-vector of size 7986 for each point

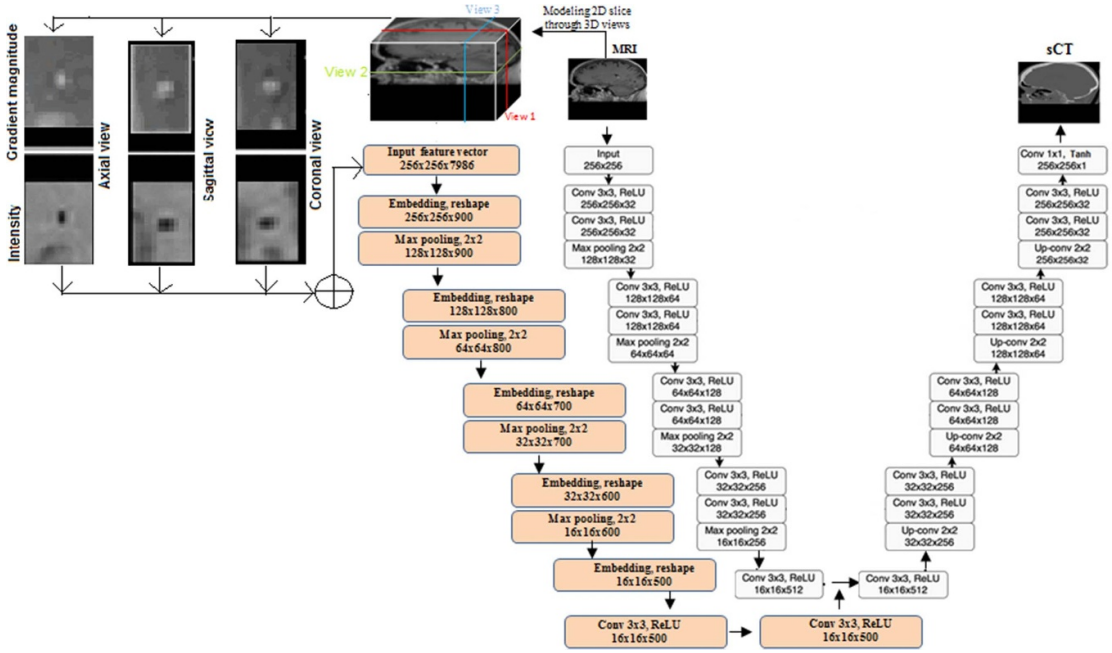


**Figure 3.** Proposed generative architecture based on dual encoding feature layer extractors using the U-Net architecture. On the left, the dual encoder branches for the 2D image (white) and the multi-planar dynamic features (orange) consistw of downsampling CNNs that generate an embedding for both feature sets before being combined via concatenation. On the right, the decoder branch that synthesizes the target sCT from the dual feature representation. For the multi-planar branch, the 2D MRI slice is first transformed in a high multi-dimensional dynamic shape feature space using a multi-planar 3D image representation. A $11 \times 11 \times 11$ patch around each point on the input 2D slice is flattened into a intensity and gradient magnitude feature vector, which is repeated for each of the three possible views (e.g. axial, sagittal and coronal). The resulting multi-planar feature descriptor consists of a feature vector of size 7986 for each point on the 2D $256 \times 256$ image, used as an input to the multi-planar branch. In each of its 5 blocks, the features are successively embedded to a lower dimensional space then downsampled by half in width and length. The final embedding is combined with the output of the 5 standard CNN blocks for the input 2D MRI, forming a $16 \times 16 \times 512$ vector as input for the transposed convolutional decoder.

on the 2D $256 \times 256$ image, that includes axial, sagittal and coronal features for both intensity and gradient magnitude. This descriptor is used as an input to the multi-planar branch. Gradient features allow to add further derivative information to guide the learned reconstruction process with enhanced spatial awareness. The gradient magnitude feature vector of the multi-planar feature representation $\nabla F$ is computed using the partial derivative along each of the two available dimensions of the considered 2D slice $x$ of the original image intensity function $F$ (Gonzalez and Woods 2018) given in equation (1)

$$F' = \text{concatenation}\left(F^{\text{axial}}, F^{\text{sagittal}}, F^{\text{coronal}}, \nabla F^{\text{axial}}, \nabla F^{\text{sagittal}}, \nabla F^{\text{coronal}}\right). \quad (2)$$

The final multi-planar dynamic feature representation $F'$ of a single 3D scan consists of the channel-wise concatenation of the intensity features $F$ along with its gradient features $\nabla F$, given in equation (2). Its shape is thus $H \times W \times C$ where $C = h * w * d * 6$. In our work, the contextual window $W$ was chosen to be of size $11 \times 11 \times 11$. This results in a final dynamic feature vector of size $256 \times 256 \times 7986$. This modeling of each 2D image of the 3D MRI volume defines a reliable and an efficient way to exploit additional spatial awareness during the transformation of the image input space domain to an output space domain in the synthesis problem.

### 3.3. Proposed model

Our proposed dual-branch GAN architecture aims to improve the 2D image synthesis problem by providing the generator model with spatial contextual information. This awareness of the localization of the input MR image is enabled by adding a secondary multi-layer neural network branch in parallel with the primary 2D MR image encoder (see figure 3). With this approach of dual feature learning, the entire model can then be trained in a fully end-to-end supervised fashion, where the translation from pixel intensity in MRI space as well as its corresponding 3D contextual multi-planar dynamic feature set are learned together.

#### *3.3.1. Dual features generator model*

First, our approach is based on the U-Net architecture composed of an encoder and a decoder path (see figure 3). The primary branch of the encoder path consists of a CNN that embeds a 2D image in the source

representation space (MRI) to a compact latent representation consisting of discriminative features for the synthesis problem. This is achieved through four convolutional blocks, each made up of two $3 \times 3$ convolutional layers with ReLU activations as well as a max-pooling layer of window size $2 \times 2$, for image downsampling. The four layers embed the learned representation into 32, 64, 128 and 256 features respectively. A final pair of $3 \times 3$ convolutional layers with ReLU activation increases the feature dimension to 512, before proceeding to the decoder path.

To improve the feature representation of the multi-planar feature vector $F'$, we encode in each block a set of two transformations: dimensionality reduction and image feature compression. We rely on a multi-planar encoder branch, which is composed of five block layers, without batch normalization. This operation is performed first via a linear recombination of the input dynamic image features vector, encoded in its channel dimension $C$, into a more compact embedding space of channel size $C'$, where $C' < C$. The projection of these feature descriptors from a higher dimensional space into a lower dimensional space aims to preserve the information relating to the structure of the spatial context of the original image, which discards the redundant information in an automated way. Following this, dynamic image features are then downsampled by half via a max-pooling layer with a window size $2 \times 2$, to compress the image itself. This entire process is repeated four times in four embedding-downsampling blocks, with a final embedding layer being followed by two $3 \times 3$ convolutional layer with a ReLU activation function following each layer. This embedding from full-sized intensity and gradient feature vector $F'$ of size $256 \times 256 \times 7986$ is performed with embedding layers with output channels 900, 800, 700, 600 and 500 respectively. The four times downsampled dynamic image feature vector $F''$ then has a shape $16 \times 16 \times 500$ before being combined with 2D convolutional branch as an input to the decoder portion of the U-Net.

The input latent representation of the 2D image combined with its compact multi-planar contextual embedding consists of a $16 \times 16 \times 512$ vector. It is processed by a set of four convolutional blocks in the decoder, mirroring the primary branch of the encoder. Blocks consists of a transposed convolutional layer with kernel size $2 \times 2$ to upsample the image size by two, followed by two $3 \times 3$ convolutional layers with ReLU activations. The decoder's feature representations through 256, 128, 64 and 32 dimensions before outputting a final $256 \times 256 \times 1$ synthesized image using the Tanh activation function in the target CT image space.

#### *3.3.2. Discriminator with high frequency pattern learning*

For the discriminator portion of our GAN architecture, a CNN based on previous works Touati *et al* (2021) for multi-scale feature extraction was adopted. This method improves the sCT generation process by augmenting the discriminator model with improved edge feature detection. Given the difference in physical image attributes between the MRI and CT modalities, boundaries were shown to be better preserved with this scheme. The discriminator network is composed of five convolutional layers. For the first four layers, we apply four convolutional filters of size $4 \times 4$, with a stride of 2 and a padding of 1. In the last layer, we use a convolutional filter of size $4 \times 4$ 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, using a scale of 0.2 and a dropout rate of 0.5. Specifically, we extract a descriptor from the learned latent representation via four quantized histograms: horizontal, vertical and both diagonals. We also extract high-intensity patterns in the embedding at each resolution as auxiliary information to augment the final output feature vector by concatenation. The number of quantized histograms for edges is: $64(H \times D/W \times W) \times 4(\Theta) \times 5(NC) = 1280$ oriented edge histograms, with $NC$ the number of convolutional layers and $\Theta$ the orientation. For each convolutional map in the discriminator network (figure 4), we divide the convolutional map into $W \times W$ square patches, then we compute the oriented quantized histograms for the left diagonal, the vertical, the right diagonal and the horizontal directions with equidistant binning of 32 bins and from the squared patch window $W \times W$. The $W \times W$ square size depends on the layer level. In the first layer, we use $32 \times 32$ square patches, and then the square window is halved by 2 for the next layer. We repeat the process until the last layer. All histograms have equidistant binning in all existing levels. 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 layer maps. The averaging score of all scales is assessed to decide if the requested synthetic sCT-image is real or fake.

### 3.4. Loss function

The training of our dual GAN model is achieved using the mixed min-max adversarial objective, where the cost function is a cross-entropy loss. Overall, it is formulated by a combination of a global loss term on both

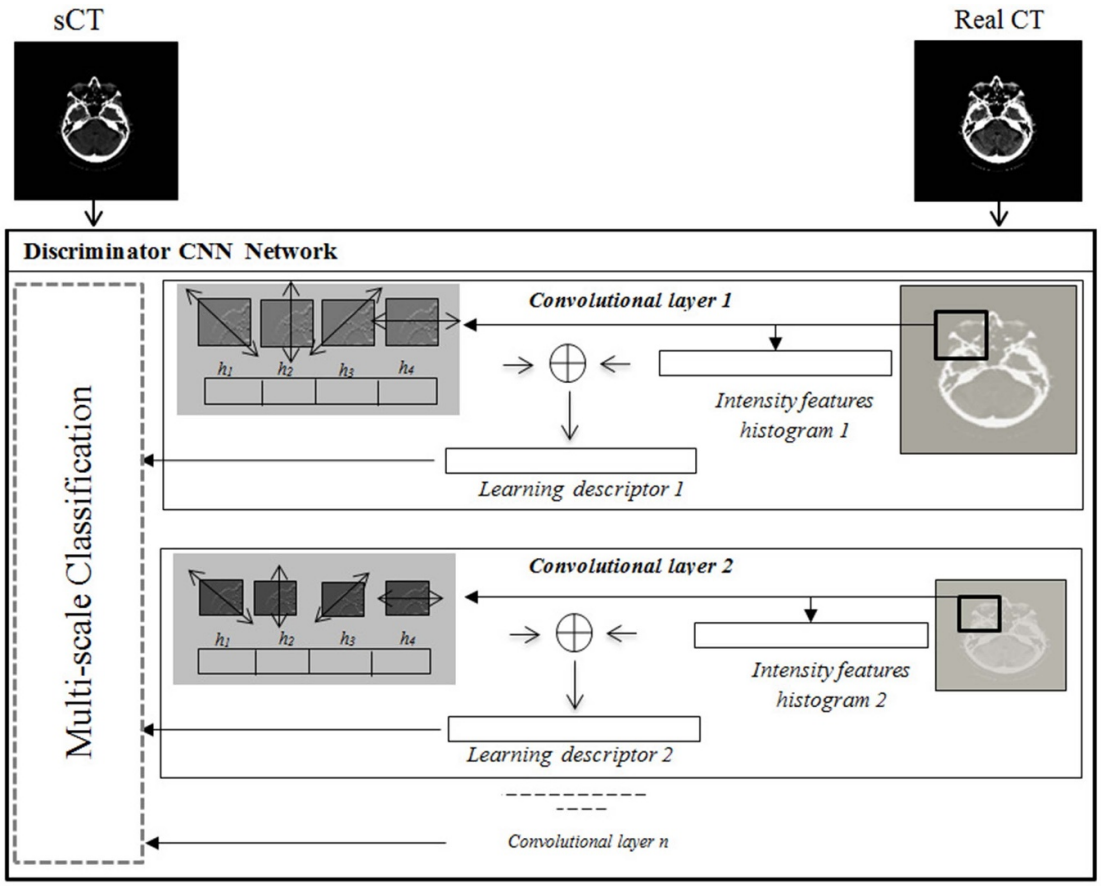


**Figure 4.** Discriminator based on high frequency and appearance patterns (Touati *et al* 2021). Our model uses the multiscale discriminator that is based on high frequency features, specifically from the boundary feature orientation. Each convolutional map is encoded with a quantized histograms $(h_1, h_2, h_3, h_4)$ describing the four edge orientations (left diagonal, the vertical, the right diagonal and the horizontal directions) and their intensity distribution. Reproduced from Touati *et al* (2021). 

the generator and the discriminator $\zeta_{\text{GAN}}(G, D)$, and of an $L_1$ term on the generator (Isola *et al* 2017). The overall min-max adversarial training cost function is given by equation (3):

$$G^* = \arg\min_G \max_D \zeta_{\text{GAN}}(G, D) + \lambda \zeta_{L_1}(G). \tag{3}$$

The generator objective function $\zeta_{\text{GAN}}(G, D)$ is given by equation (4):

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

where x is the input MRI and y is the target CT. The $L_1$ term $\zeta_{L_1}(G)$ is defined in equation (5), with a loss weight $\lambda$:

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

From a given observed (MRI) image, our dual CT-synthesis GAN model learns to generate an output mapping of the CT image $y$. The generator network $G$ synthesizes from the noise vector $x$, using the proposed dual feature representation, an image that follows the unknown CT image probability distribution $p_y$. The discriminator $D$ classifies the generated $y$ image in order to distinguish the real CT image from the synthetically generated sCT. The model is learned by an alternating training scheme, once for the generator $G$ and once for the discriminator $D$ networks. The generator network $G$ minimizes the expected objective loss while the discriminator $D$ maximizes it, where the learned features is used to compute the loss function.

### 3.5. Training and evaluation

In this work, two datasets were used for training and evaluation purposes: DS1, which is comprised of sagittal plane acquisitions of head and neck scans used to synthesize CT for model validation; and DS2, which comprised of axial plane acquisitions and was used for training and testing of head and neck scans with clinically confirmed tumors via expert segmentations. A cross-validation procedure on the sagittal acquisitions dataset DS1 was performed using 5 folds with 180 training samples and 45 validation samples. The validation with DS2 was performed on the axial acquisitions dataset DS2 with 200 samples used for training and 41 tumor samples used for testing. In all cases, online data augmentation techniques were used, including random cropping ($256 \times 256$) and random flipping, each with 50% independent probability.

In order to assess the quality of the generated sCT image of the synthesis models, we propose to evaluate the generated images based on two different aspects: global image metrics (Touati *et al* 2021) and tumor-local metrics. For global image similarity between sCT and ground truth CT, mean absolute error (MAE) and peak-signal-to-noise-ratio (PSNR) metrics were utilized to evaluate the HU intensity space, which is important to preserve clinically relevant electron density information. The contrast and the anatomy were evaluated using the mean structural similarity (MSSIM) (Dong *et al* 2017), Pearson cross-correlation (PCC) coefficient (Lauritzen *et al* 2019), Frechet inception distance (FID) (Heusel *et al* 2017), and sliced Wasserstein distance (SWD) (Deshpande *et al* 2018) measures. We also evaluated the shape quality of the different HU thresholded areas of the sCT image using the Dice score (Crum *et al* 2006, Milletari *et al* 2016), where the sCT image is segmented into to three different regions corresponding to bone ($HU > 300$), air ($HU < -100$) and soft tissues ($-100 < HU < 300$). Finally, we assessed the histogram intersection quality using the Bhattacharyya distance (BD) (Kailath 1967).

For tumor-local evaluation, the same image metrics as described above, as well as histogram similarity metrics were used on the segmented region, as well as a Dice score to measure the overlap between the synthesized tumor masks and the expert segmentation. This enables to compare the results with the ground truth tumor segmentation on the original CT.

In the sets of experiments, we consider the following set of optimized parameters values, as determined in a previous study on the feature GAN (Oulbacha and Kadoury 2020, Touati *et al* 2021): a momentum equal to 0.5, epoch number equal to 600, learning rate initialized to 0.0002 then reduced linearly to 0 starting at epoch 200, L1 weight $\lambda = 100$, batch size of 1, and Adam optimizer (Kingma and Jimmy 2015). The number of bins used for the quantized histogram in the discriminator is 32.

## 4. Results

In our experiment, we computed different imaging evaluation metrics used for CT/MR image synthesis evaluation (Touati *et al* 2021, Touati and Kadoury 2023). We computed the MAE and PSNR that consider the intensity difference, while we computed the MSSIM index to evaluate the contrast level of the sCT. The FID evaluates the similarity between the sCT and the real CT using overall statistics, while the SWD reports the overall deviation between the sCT and CT. The BD evaluates the histogram quality distribution between the sCT and CT intensity spaces, while the Dice score evaluates the overlap between normal or tumor segmentations of sCT and CT.

### 4.1. Ablation study

We present in table 1 an ablation study for generating CT images from the T1 input, using our model with and without the intensity as well as gradient features, the addition of quantized histograms in the discriminator, as well as assessing the loss function based on SSIM and BD metrics, used in these experiments to replace the L1 generator component ($\zeta_{L1}$). In both cases, the hyperparameter $\lambda$ was maintained at 100. We report the performance of our method using one plane (axial) and two planes (axial/sagittal). We also compare our method to the baseline 3D U-net. The evaluation was performed on DS1, using 5 folds cross validation procedure.

We finally experimented the model for generating axial CT for DS2, using the model trained only on the sagittal images from DS1. The following set of experiments trained the model on mixed training data that combines sagittal and axial images from the extended DS1 and DS2 datasets, using both sagittal and axial images. The dataset was split such that the training set included 475 unannotated cases, and the testing set included 63 tumor segmented cases. The trained model was then assessed using the tumor test set, for generating axial tumor sCT from MRI-T1. Quantitative results of the two experiments are presented in table 7. We can observe that the model fails when attempting to generate axial sCT using the model trained on the sagittal DS1 (table 2 and figure 5). This can be explained by the fact that the DS1 and DS2 datasets originate from independent acquisitions with different orientations, which requires a more complex model to extract robust features that can generalize from one orientation to another. Furthermore, the model is not

**Table 1.** Ablation study.

| Method | MAE (std) Hu |
|---|---|
| Without intensity | 134 (7.624) |
| Without gradient | 118 (8.204) |
| One plane | 127.23 (6.325) |
| Two planes | 104.15 (6.451) |
| Without quantized histograms | 70.50 (6.325) |
| 3D U-net | 61.83 (7.403) |
| SSIM metric loss | 32.48 (5.124) |
| BD metric loss | 30.69 (6.163) |
| Proposed | 18.76 (5.167) |

**Table 2.** Model performance for (1) generating axial sCT of DS2 test set, using the trained model on sagittal DS1, and (2) generating axial tumor sCT using a combined training DS1 and DS2.

| Training dataset | Test dataset | MAE (std) HU |
|---|---|---|
| DS1 (sagittal) | DS2 (axial) | 452.14 (11.13) |
| Combined DS1 + DS2 | DS2 (63 tumor cases) | 69.85 (7.20) |
| Combined DS1 + DS2 | DS1 | 48.12 (6.50) |

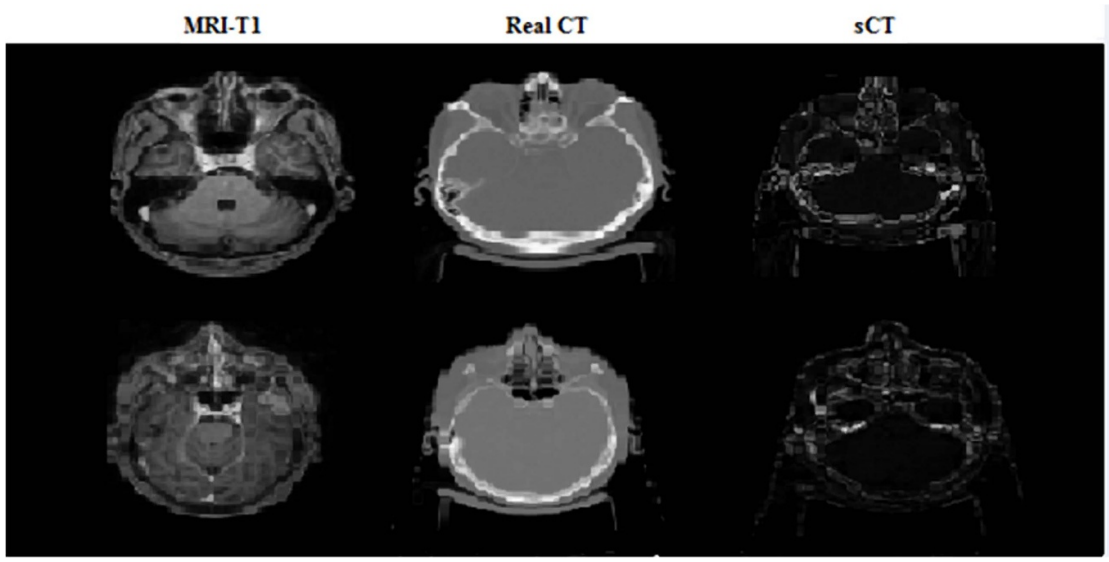


**Figure 5.** Generation of the axial sCT images from the DS2 test set using the model trained only on the sagittal DS1 dataset.

able to generate images with completely different orientation planes, such as producing axial planes when using only sagittal planes during training, due to the fact the two datasets DS1 (sagittal) and DS2 (axial) were acquired separately and of a completely different nature with distinct physical acquisition properties and varying resolution fields of view. This significant bias between the datasets and significant varying acquisitions protocols, along with limited size of the datasets, hindered the model's generalization capability.

On the other hand, we can see that the model remains flexible to generate axial tumor sCT, when using the model trained from the combined DS1 and DS2 datasets (table 2 and figure 6). In such a case, we believe that the combination of axial and sagittal acquisitions helps in the generalizability of the model, which is guided through multi-planar training features and from a combined observation of both DS1 and DS2. We also performed an experiment to evaluate the model trained on the pooled DS1+DS2 dataset, using mixed acquisition inputs (sagittal for DS1, axial for DS2) but evaluated on a held out subset of DS1 (50 cases) using sagittal inputs only, as shown in table 2.

### 4.2. Comparison with state-of-the-art methods

*4.2.1. Cross-validation*

In the next set of experiments, we evaluated the models on DS1, with the performance assessed on whole sCT images. Table 3 presents the obtained results using the 5-fold cross-validation for the different imaging

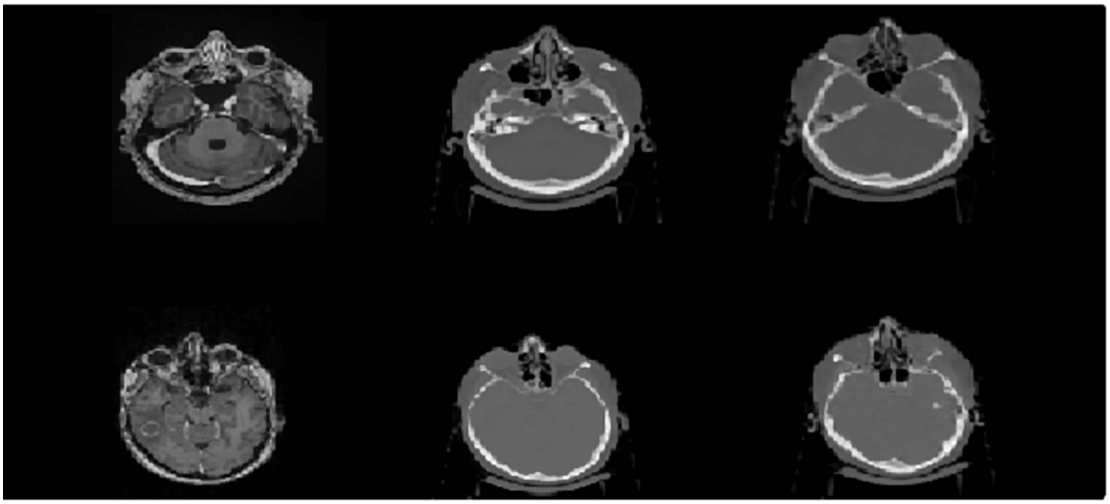


**Figure 6.** Generation of the axial sCT images from tumor cases, using the trained model from combined DS1 and DS2 datasets (sagittal and axial acquisitions).

**Table 3.** CT synthesis evaluation on DS1 comparing the proposed model with three state-of-the-art approaches. The generated image quality is indicated by the ⇑ and ⇓ which determine respectively that highest and lowest values leading to a better image generation performance in the result comparison.

| | MAE (⇓) | PSNR (⇑) | MSSIM (⇑) | PCC (⇑) |
|---|---|---|---|---|
| Cycle GAN | 213.45 (7.45) | 22.80 (0.31) | 0.75 (0.01) | 0.81 (0.04) |
| Conditional GAN | 109.53 (8.54) | 24.10 (0.66) | 0.79 (0.04) | 0.87 (0.03) |
| Feature GAN | 64.50 (5.28) | 28.30 (0.74) | 0.85 (0.07) | 0.90 (0.06) |
| Proposed GAN | 18.76 (5.17) | 33.90 (0.46) | 0.95 (0.09) | 0.95 (0.02) |
| | $\text{Dice}_{\text{bone}}$ (⇑) | $\text{Dice}_{\text{soft}-\text{tissue}}$ (⇑) | | $\text{Dice}_{\text{air}}$ (⇑) |
| Cycle GAN | 0.64 (0.03) | 0.73 (0.04) | | 0.77 (0.01) |
| Conditional GAN | 0.66 (0.02) | 0.75 (0.07) | | 0.78 (0.05) |
| Feature GAN | 0.68 (0.06) | 0.77 (0.04) | | 0.80 (0.03) |
| Proposed GAN | 0.75 (0.05) | 0.82 (0.09) | | 0.84 (0.01) |
| | FID (⇓) | SWD (⇓) | | BD (⇑) |
| Cycle GAN | 432.20 (5.75) | 55.09 (10.99) | | 0.78 (0.12) |
| Conditional GAN | 314.70 (7.86) | 48.16 (11.82) | | 0.80 (0.10) |
| Feature GAN | 269.34 (9.50) | 37.57 (7.42) | | 0.85 (0.09) |
| Proposed GAN | 145.60 (8.38) | 16.41 (3.52) | | 0.91 (0.04) |

metrics as well as image overlap of the segmented tissue regions by HU thresholds. Results show that the proposed model gives the best average MAE of 18.76(5.17), contrast enhancement FID of 145.60(8.39) and MSSIM of 0.95(0.09), and shape preservation Dice score of 0.75(0.05), 0.82(0.09) and 0.84(0.01) for bone, soft-tissues and air regions respectively. We report a better BD score of 0.91(0.04) for histogram intensity distribution intersection. Figure 7 shows a qualitative comparison results of four patients, where we can observe that our proposed model produces an improved sCT image compared to other synthesis techniques. Also, the comparison of the histogram distributions shows that the intensity distribution of our generated sCT image match the intensity distribution of the real sCT image, contrary to the other models. We also reported a comparative study between our model and the HMSS-Net (Li *et al* 2023) method using the extended DS1 and DS2 datasets. Table 4 reports the quantitative comparison results in the HU space. Compared to the HMSS-Net (Li *et al* 2023) model, our model achieves improved synthesis results.

To assess the performance of our model on MRI-T1 generation from the CT space domain, we report in table 5 the obtained MRI-T1 synthesis results of the proposed and the state-of-the-art models. We also show in figure 8(a) qualitative comparison for the different models on sagittal slices DS1. The presented MRI-T1 synthesis results show that our dual feature model reports a lower MAE of 13.456(5.842) and a better synthesis MRI-T1 image quality, contrary to other synthesis models.

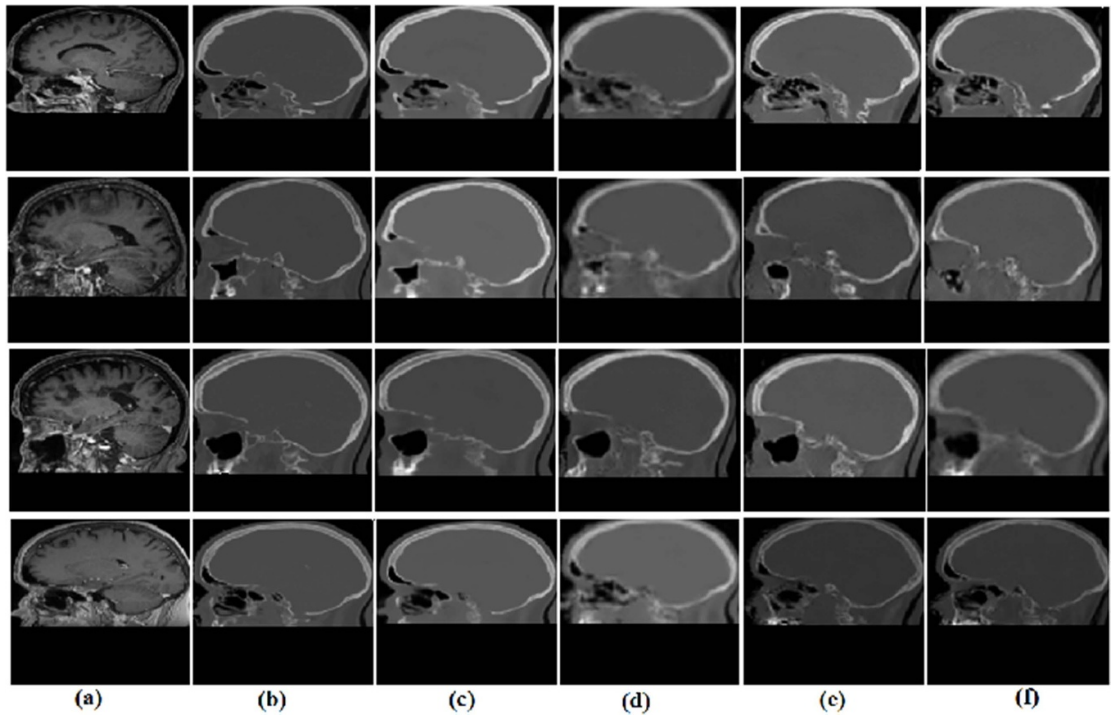


**Figure 7.** Synthetic CT comparison between the proposed model and state-of-the-art approaches on DS1. (a) Input MRI-T1 image, (b) Ground truth CT, (c) Proposed model, (d) Feature GAN, (e) Conditional GAN, (f) Cycle GAN.

**Table 4.** Comparison between the proposed method and the HMSS-Net (Li *et al* 2023) in the Hu space, using the DS1 and DS2 datasets.

| Dataset | Method | MAE (std) HU |
|---|---|---|
| DS1 | HMSS-Net (Li *et al* 2023) | 65.02 (5.801) |
| | Proposed | 37.45 (6.215) |
| DS2 | HMSS-Net (Li *et al* 2023) | 74.49 (7.523) |
| | Proposed | 43.25 (7.107) |

**Table 5.** MRI-T1 synthesis evaluation obtained on DS1 by the proposed model compared with three state-of-the-art approaches. The generated image quality is indicated by the ⇑ and ⇓ which determine respectively that highest and lowest values leading to a better image generation performance in the result comparison.

| | MAE (⇓) | PSNR (⇑) | MSSIM (⇑) | PCC (⇑) |
|---|---|---|---|---|
| Cycle GAN | 117.07 (6.98) | 16.81 (2.47) | 0.66 (0.07) | 0.70 (0.03) |
| Conditional GAN | 57.11 (5.21) | 19.08 (2.16) | 0.75 (0.02) | 0.76 (0.15) |
| Feature GAN | 44.61 (6.27) | 22.44 (2.412) | 0.82 (0.09) | 0.84 (0.09) |
| Proposed GAN | 13.46 (5.84) | 27.89 (2.22) | 0.89 (0.12) | 0.91 (0.03) |
| | FID (⇓) | SWD (⇓) | | BD (⇑) |
| Cycle GAN | 234.01 (3.11) | 30.84 (3.23) | | 0.68 (0.03) |
| Conditional GAN | 175.44 (7.89) | 27.89 (2.67) | | 0.74 (0.05) |
| Feature GAN | 138.94 (5.57) | 18.26 (3.88) | | 0.81 (0.10) |
| Proposed GAN | 94.80 (3.32) | 10.54 (2.71) | | 0.87 (0.08) |

*4.2.2. Head and neck tumor evaluation*

In this set of experiments, we assessed the synthesis performance of the proposed model compared with state-of-the-art approaches on the tumor-local region for DS2. Table 6 presents the quantitative results. We can observe from the obtained results that our proposed model achieves a lower MAE error of 26.83(8.27) compared to the other models, as well as a higher Dice score of 0.73(0.06) and a better BD score of 0.85(0.09). We present in figure 9(a) visual comparison of synthesized tumor regions (tumor regions appears

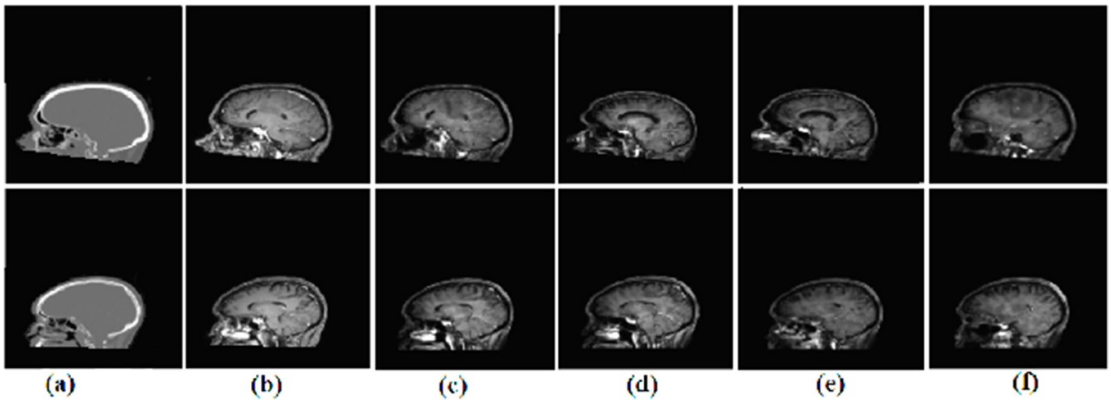


**Figure 8.** Synthesis of MRI-T1 comparison between the proposed model and state-of-the-art approaches on DS1. (a) Input CT image, (b) Ground truth MRI-T1, (c) Proposed model, (d) Feature GAN, (e) Conditional GAN, (f) Cycle GAN.

**Table 6.** CT synthesis evaluation on DS2 comparing the proposed model with three state-of-the-art approaches. The quality of the generated tumor regions is indicated by the ⇑ and ⇓ which determine respectively that highest and lowest values leading to a better image generation performance in the result comparison. $\text{Dice}_{\text{GTV}}$ is the dice score quantifying the shape intersection quality between two segmented gross tumour volumes (GTV).

| | MAE (⇓) | PSNR (⇑) | MSSIM (⇑) | PCC (⇑) |
|---|---|---|---|---|
| Cycle GAN | 276.43 (9.31) | 15.33 (3.11) | 0.79 (0.04) | 0.81 (0.03) |
| Conditional GAN | 160.53 (8.54) | 17.31 (3.66) | 0.81 (0.02) | 0.86 (0.04) |
| Feature GAN | 84.93 (7.01) | 25.37 (2.75) | 0.87 (0.07) | 0.91 (0.02) |
| Proposed GAN | 26.83 (8.27) | 29.14 (4.58) | 0.94 (0.05) | 0.95 (0.03) |
| | $\text{Dice}_{\text{GTV}}$ (⇑) | FID (⇓) | SWD (⇓) | BD (⇑) |
| Cycle GAN | 0.56 (0.09) | 332.42 (5.55) | 45.09 (8.99) | 0.68 (0.19) |
| Conditional GAN | 0.62 (0.03) | 217.94 (9.76) | 34.16 (8.43) | 0.71 (0.11) |
| Feature GAN | 0.67 (0.02) | 179.08 (6.35) | 22.57 (7.44) | 0.78 (0.22) |
| Proposed GAN | 0.73 (0.06) | 122.58 (7.55) | 12.41 (5.53) | 0.85 (0.19) |

in red color), which shows that the tumor region in the sCT image is the most similar to the tumor in the ground truth image in terms of shape and size compared to the others CTs images.

We have also quantified in table 7 the performance of the proposed models on test images for generating axial MRI-T1 image from the input CT. The obtained results show that our model yields significant improvements over the MAE compared to the other models with an error value of 21.21(6.07). The qualitative results presented in figure 10 also confirm the synthesis ability of our model where the organs appear similar to those in the ground truth MRI image. Moreover, our model produced MRI images with improved accuracy compared to other methods, where the overall sMRI images better SSIM and SNR metrics compared to the others. This can be explained by the synthesis ability of our model that incorporates an augmented multi-planar branch that captures more global and local features through the three views. This is possible because of the fact we used paired CT-MRI training data, where this purely data-driven approach will exploit rich anatomical information which is enhanced between the modalities.

## 5. Discussion

Through the qualitative and quantitative experimental results, it can be observed that the 2D conditional, 2D feature GANs and the 3D CycleGAN results in lower performance compared to our proposed model on both sagittal DS1 and axial DS2 head-and-neck datasets, both for generating the whole sCT and for synthesizing specific primary tumor regions. We achieved a Dice scores of 0.75(0.05), 0.82(0.09), 0.84(0.01), on the segmented components such as bone, soft tissue, and air areas, respectively, which indicates a good generation quality in the different components. This is explained by the ability of the proposed feature representation leading to improved qualitative results in different heterogeneous area. In addition, our model yields the highest Dice score (0.73(0.06)) for the overlapping tumor regions, demonstrating the model's capability of preserving specific subregions of interests, such as the GTV. This is explained by the integration of the dynamic shape features in the training process of the generator. The model is able not only to preserve the structures, similar to how the feature GAN (Touati *et al* 2021) does by integrating the information of the

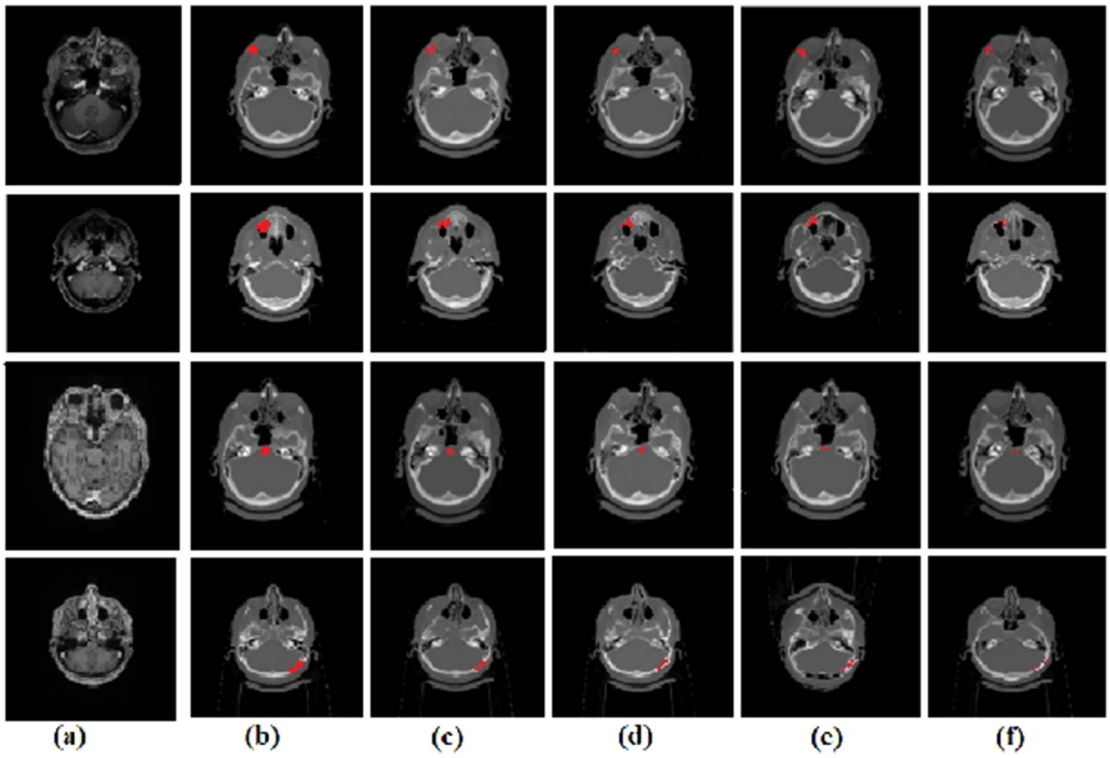


**Figure 9.** Synthetic CT comparison between the proposed model and state-of-the-art approaches on DS2. The red regions represent the correctly synthesized tumor region, that correspond to the intersection between the correctly synthesized cancer region and the segmented cancer regions in the ground truth: (a) Input MRI-T1 image, (b) Ground truth of CT and cancerous region appearing in red (superimposed in red) on the corresponding CT image, (c) Proposed model, (d) Feature GAN, (e) Conditional GAN, (f) Cycle GAN.

**Table 7.** MRI-T1 synthesis evaluation obtained by our model on the tumor axial dataset (DS2), with different synthesis models based on different learning network architectures. The generated image quality is indicated by the (⇑) and (⇓) which determine respectively that highest and lowest values giving the better image generation performance in the result comparison.

| | MAE (⇓) | PSNR (⇑) | MSSIM (⇑) | PCC (⇑) |
|---|---|---|---|---|
| Cycle GAN | 196.54 (4.02) | 15.41 (2.01) | 0.65 (0.03) | 0.66 (0.05) |
| Conditional GAN | 86.20 (5.02) | 18.99 (1.98) | 0.71 (0.04) | 0.73 (0.01) |
| Feature GAN | 72.63 (4.91) | 20.38 (1.91) | 0.80 (0.09) | 0.82 (0.03) |
| Proposed GAN | 21.21 (6.07) | 26.08 (2.95) | 0.86 (0.08) | 0.88 (0.05) |
| | FID (⇓) | SWD (⇓) | | BD (⇑) |
| Cycle GAN | 294.13 (5.51) | 32.83 (2.49) | | 0.63 (0.04) |
| Conditional GAN | 205.52 (4.42) | 30.42 (3.06) | | 0.72 (0.08) |
| Feature GAN | 178.21 (7.79) | 20.63 (2.12) | | 0.80 (0.05) |
| Proposed GAN | 127.48 (9.09) | 14.00 (3.15) | | 0.85 (0.09) |

contours, but is also able to generate regions most similar to the real CT in terms of intensity distribution and anatomical characteristics.

In all experiments, the performance of our model shows that it does not lose flexibility on MRI-T1 synthesis for both sagittal and axial acquisition head-and-neck datasets. We note however that the obtained MRI intensity distribution differs slightly from the MRI ground truth due to the slight degradation in the MRI synthesis process: the higher inherent variety of physical scanning properties and texture complexity in MRI due to absence of a standard intensity unit (such as is the case for HU in CT) leads us to postulate that a larger dataset may be required to truly cover the image distribution in the sMRI space. Nonetheless, this demonstrates that our model can be also used for different input/output intensity space pairs for image synthesis.

Other state-of-the-art CT synthesis models evaluated in this study, decode the sCT image from the latent encoding space, where the latest encoding feature maps represent only the spatial convoluted features or the convolutional maps generated in terms of high frequency patterns (Touati *et al* 2021). Our dual model on the other hand considers both spatial and multi-planar features and transforms them in a relevant dual encoding

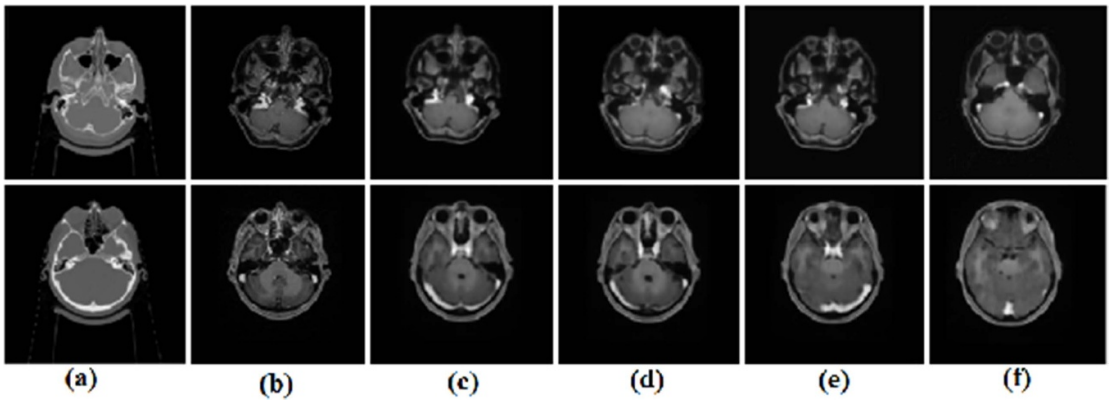


**Figure 10.** Synthesis of MRI-T1 comparison between the proposed model and state-of-the-art approaches on DS2. (a) Input CT image, (b) Ground truth MRI-T1, (c) Proposed model, (d) Feature GAN, (e) Conditional GAN, (f) Cycle GAN.

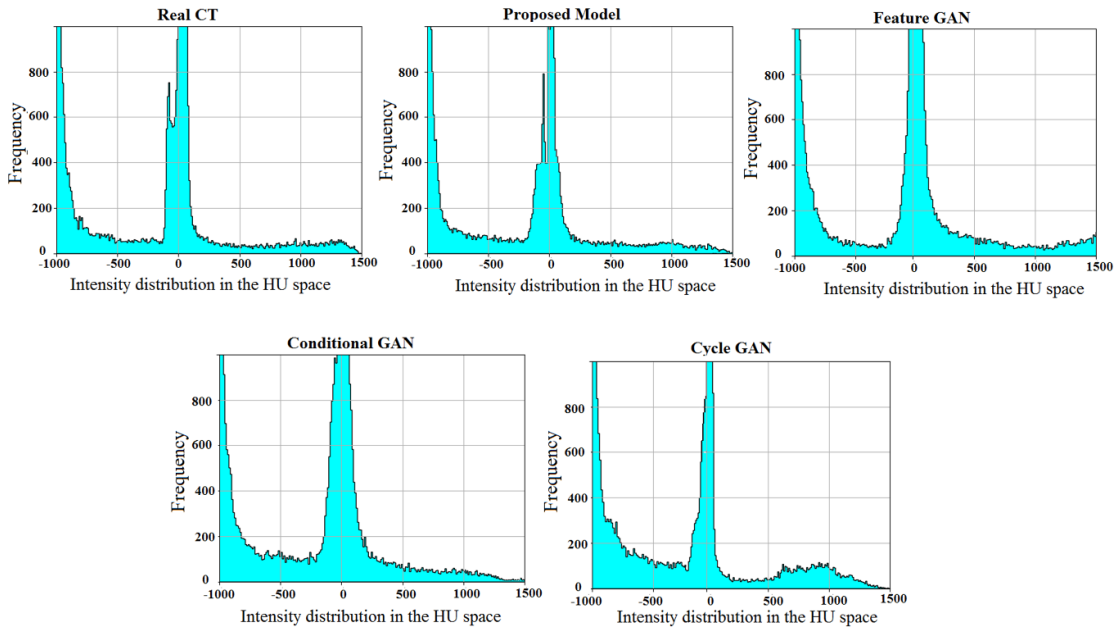


**Figure 11.** CT intensity histogram comparison between the different obtained sCT images in the (HU) space.

latent feature space due in part to our learning based feature extraction strategy. Indeed, our model also considers the transformation of the high-dimensional shape feature of the MRI in a more relevant encoding of CT feature shape maps thanks to its embedding layers. Thus, our model architecture is able to learn a combination of multi-planar features that increase the synthesis performance. This suggests that our deep synthesis model based on dual embedding and convolutional layers takes into account the relevant complementary information between the 2D oriented views of the full volumetric MRI, as well as the convolutional feature maps. This indicates that these feature maps capture not only the convolution features, but also the shape feature maps resulting from the learned embedding based multi-planar strategy. This allows the proposed model to improve synthesis as opposed to comparative methods, especially in high resolution region that contains small details of the anatomies such us the tumor regions where the regions shapes vary highly (see figures 9, 11 and 12).

In the context of glioblastoma segmentation in brain images, the obtained results are considered to be acceptable, giving the fact this is an extremely challenging task and is performed on CT images. The literature reports an inter-rater variability from manual annotations between 0.75 and 0.80 (Porz *et al* 2014). In future work, we will collect clinical measurements and evaluate the model using dose plan parameters for radiotherapy planning. We will conduct experiments on mixed imaging acquisitions (coronal, sagittal and axial).

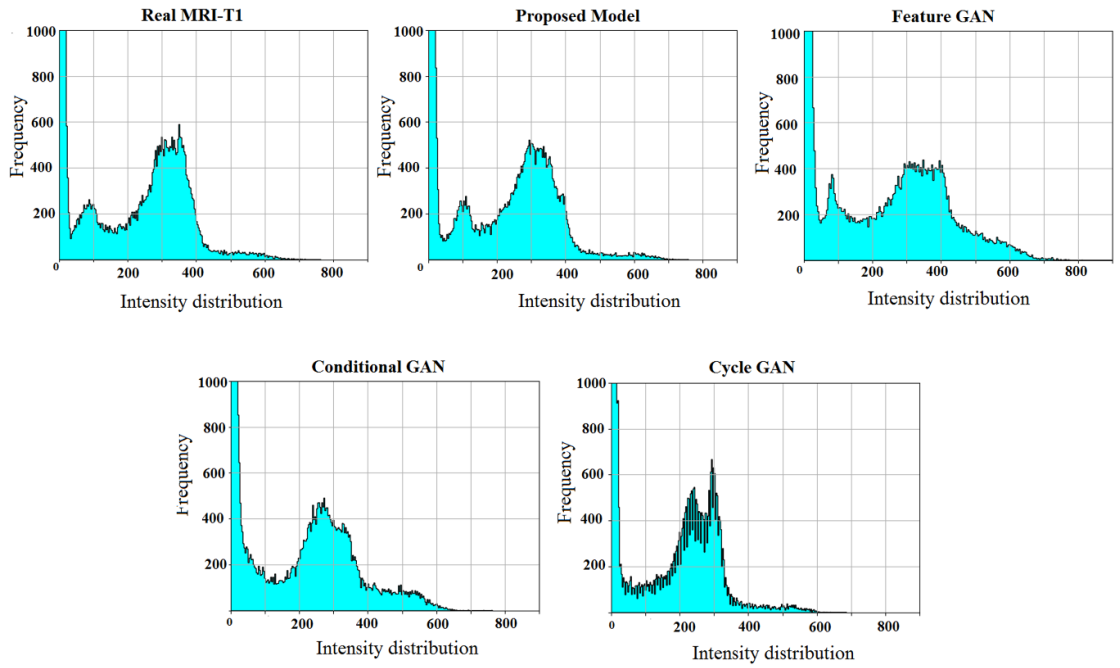


**Figure 12.** MRI-T1 intensity histogram comparison between the different generated MRI images.

Our study was validated on head and neck cancer datasets covering two different plane acquisitions, where collecting the corresponding dose plan parameters for dosimetric evaluation remains to be explored. Additionally, we can not assume that our model will generalize better to other body anatomies since our validation was performed only on head and neck anatomy. In future works, we will validate the proposed approach considering a larger dataset from different clinical applications, involving the generation of dose plans on the synthetic sCT images from MRI acquisitions, where we intend also to improve the MRI-T1 quality synthesis. We plan to demonstrate the generalization capability of our model to synthesize sCT image using a combination of multiple MRI sequences.

Limitations to this study are the limited sizes of the datasets, which can be complemented with other publicly available sources, as well as the lack of homogeneity in the data acquisition parameters. Furthermore, the model does not exploit full 3D feature extraction capabilities, which could potentially improve the generalization of the model to other types of sequences. In addition, the model is trained on head and neck acquisitions with a fixed resolution, which can be improved using a variety input resolutions. Also, the model was validated for generating sCT from MRI-T1 inputs, in which the model is less effective to generate sCT from noisy MRI-T1 input or other highly noisy MRI contrasts. The model is also validated under paired-training data. To address this issue, we plan to augment the training set used by the model with unpaired training data, using self-supervised training strategy. Finally, our model can be extended to include a cycle learning strategy in order to further enhance the quality of the generated MRI image.

## 6. Conclusion

In this paper, we presented a novel MRI-to-CT image generation model for head and neck cancer radiotherapy applications. In the generation phase, our proposed dual synthesis GAN model learns multi-planar image dual dynamic features that are extracted not only from the input 2D image but also from extracted multi-planar images that exploits the different image views of the full MRI-T1 volume, by using a strategy based on the convolutional U-net network with two parallel encoder branches. This end-to-end learning strategy allows us to generate an improved representation of the sCT image from the dual latent spaces in the decoder, thanks to the combination of 2D CNN features as well as multi-planar intensity and gradient shape embedding features. The conducted experiments on two different datasets confirm the efficiency of our MRI-to-CT synthesis in different image acquisition planes—sagittal and axial—and the robustness of the model to generate specific tumor regions in head and neck T1-weighted MRI datasets with different spatial resolutions.

## Data availability statement

The data cannot be made publicly available upon publication due to legal restrictions preventing unrestricted public distribution. The data that support the findings of this study are available upon reasonable request from the authors.

## Acknowledgments

This work is supported by the funding agencies NSERC (GPIN-2020-06558) and FRQS (293740).

## 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. This study was performed on two public human brain subject datasets. Ethical approval was confirmed by the licence attached with their open access data.

## ORCID iDs

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

## References

Abbasian Ardakani A, Alireza Rajabzadeh Kanafi U R A, Khadem N and Mohammadi A 2020 Application of deep learning technique to manage covid-19 in routine clinical practice using ct images: Results of 10 convolutional neural networks *Comput. Biol. Med.* **121** 103795

Abu-Srhan A, Almallahi I, Abushariah M A M, Mahafza W and Al-Kadi O S 2021 Paired-unpaired unsupervised attention guided gan with transfer learning for bidirectional brain mr-ct synthesis *Comput. Biol. Med.* **136** 104763

Askin B, Güngör A, Alptekin Soydan D, Ulku Saritas E, Barış Top C and Cukur T 2022 Pp-mpi: A Deep Plug-and-Play Prior for Magnetic Particle Imaging Reconstruction *Int. Workshop on Machine Learning for Medical Image Reconstruction* (Springer) pp 105–14

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

Dalmaz O, Yurt M and Cukur T 2022 Resvit: Residual vision transformers for multimodal medical image synthesis *IEEE Trans. Med. Imaging* **41** 2598–614

Deshpande I, Zhang Z and Schwing A G 2018 Generative modeling using the sliced wasserstein distance *Proc. IEEE Conference on Computer Vision and Pattern Recognition* pp 3483–91

Dinkla A M, Florkow M C, Maspero M, Savenije M H F, Zijlstra F, Doornaert P A H, van Stralen M, Philippens M E P, van den Berg C A T and Seevinck P R 2019 Dosimetric evaluation of synthetic ct for head and neck radiotherapy generated by a patch-based three-dimensional convolutional neural network *Med. Phys.* **46** 4095–104

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 *Medical Image Computing and Computer Assisted Intervention, MICCAI 2017* (Springer International Publishing) 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–e11

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 (Pearson)

Goodfellow I, Bengio Y and Courville A 2016 *Deep Learning* (MIT Press)

Goodfellow I, Pouget-Abadie J, Mirza M, Bing X, Warde-Farley D, Ozair S, Courville A and Bengio Y 2014 Generative adversarial nets *Advances in Neural Information Processing Systems* pp 2672–80

Han X 2017 Mr-based synthetic ct generation using a deep convolutional neural network method *Med. Phys.* **44** 1408–19

Hattangadi J A *et al* 2012 Single fraction proton beam stereotactic radiosurgery (psrs) for inoperable cerebral arteriovenous malformations (avms) *Int. J. Radiat. Oncol. Biol. Phys.* **84** S38

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 *Advances in Neural Information Processing Systems* pp 6626–37

Hsu S-H, Han Z, Leeman J E, Hu Y-H, Mak R H and Sudhyadhom A 2022 Synthetic ct generation for mri-guided adaptive radiotherapy in prostate cancer *Front. Oncol.* **12** 969463

Huynh T, Gao Y, Kang J, Wang Li, 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

Işın A, Direkoğlu C and Şah M 2016 Review of mri-based brain tumor image segmentation using deep learning methods *Proc. 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 Thomas Asklund and Tufve Nyholm. 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 *et al* 2020 Dosimetric evaluation of synthetic ct generated with gans for mri-only proton therapy treatment planning of brain tumors *J. Appl. Clin. Med. Phys.* **21** 76–86
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 Jimmy B 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 *Med. Phys.* (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)* pp 1–6
Li Y, Xu S, Lu Y and Qi Z 2023 Ct synthesis from mri with an improved multi-scale learning network *Front. Phys.* **11** 1088899
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, Ehteshami Bejnordi B E, Arindra Adiyoso Setio A, Ciompi F, Ghafoorian M, Awm Van Der Laak J, Van Ginneken B and Sánchez C I 2017 A survey on deep learning in medical image analysis *Med. Image Anal.* **42** 60–88
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.* **153** 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 International Conference on 3D Vision (3DV)* (IEEE) pp 565–71
Narin A 2021 Accurate detection of covid-19 using deep features based on x-ray images and feature selection methods *Comput. Biol. Med.* **137** 104771
Nie D, Cao X, Gao Y, Wang Li and Shen D 2016 Estimating ct image from mri data using 3d fully convolutional networks *Deep Learning and Data Labeling for Medical Applications* (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* (Springer) pp 417–25
Ninon Burgos M J C, Guerreiro F, Veiga C, Modat M, McClelland J, Knopf A-C, Punwani S, Atkinson D and Arridge S R *et al* 2015 Robust ct synthesis for radiotherapy planning: application to the head and neck region *Int. Conf. on Medical Image Computing and Computer-Assisted Intervention* (Springer) pp 476–84
Oulbacha R and Kadoury S 2020 MRI to CT synthesis of the lumbar spine from a pseudo-3d cycle GAN *17th IEEE Int. Symp. on Biomedical Imaging, ISBI 2020 (Iowa City, IA, USA, 3–7 April 2020)* (IEEE) pp 1784–7
Ozbey M, Dalmaz O, Dar S U H, Bedel H A, Ozturk Ş, Gungor A and Cukur T 2023 Unsupervised medical image translation with adversarial diffusion models *IEEE Trans. Med. Imaging* **42** 3524–39
Ozturk T, Talo M, Azra Yildirim E A, Baran Baloglu U B, Yildirim O and Rajendra Acharyaf U 2020 Automated detection of covid-19 cases using deep neural networks with x-ray images *Comput. Biol. Med.* **121** 103792
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
Porz N, Bauer S, Pica A, Schucht P, Beck J, Kumar Verma R K, Slotboom J, Reyes M, Wiest R and Strack S 2014 Multi-modal glioblastoma segmentation: man versus machine *PLoS One* **9** e96873
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 *Jpn. J. Clin. Oncol.* **37** 405–11
Purdy J A, Perez C A and Poortmans P 2012 *Technical Basis of Radiation Therapy: Practical Clinical Applications* (Springer)
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
Redha Touati and Samuel Kadoury 2023 Bidirectional feature matching based on deep pairwise contrastive learning for multiparametric mri image synthesis *Phys. Med. Biol.* **68** 125010
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) (Lncs)* vol 9351 (Springer) pp 234–41
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
Touati R and Kadoury S 2023 A least square generative network based on invariant contrastive feature pair learning for multimodal mr image synthesis *Int. J. Comput. Assisted Radiol. Surgery* **18** 971–9

Touati R, Le W T and Kadoury S 2021 A feature invariant generative adversarial network for head and neck mri/ct image synthesis *Phys. Med. Biol.* **66** 095001

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 *Int. Workshop on Simulation and Synthesis in Medical Imaging* (Springer) pp 14–23

Yurt M, Dar S U, Erdem A, Erdem E, Erdem E, Oguz K K and Cediiukur T 2021 mustgan: multi-stream generative adversarial networks for mr image synthesis *Medical Image Analysis* **70** 101944

Zhu J-Y, Park T, Isola P and Efros A A 2017 Unpaired image-to-image translation using cycle-consistent adversarial networks *Proc. IEEE Int. Conf. on Computer Vision* pp 2223–32

Zili Y, Zhang H, Tan P and Gong M 2017 Dualgan: Unsupervised dual learning for image-to-image translation *Proc. IEEE Int. Conf. on computer vision* 2849–57