

Article

Comparison of Supervised and Unsupervised Approaches for the Generation of Synthetic CT from Cone-Beam CT

Matteo Rossi *  and Pietro Cerveri * 

Department of Electronics, Information and Bioengineering, Politecnico di Milano, 20133 Milan, Italy

* Correspondence: matteo2.rossi@polimi.it (M.R.); pietro.cerveri@polimi.it (P.C.)

Abstract: Due to major artifacts and uncalibrated Hounsfield units (HU), cone-beam computed tomography (CBCT) cannot be used readily for diagnostics and therapy planning purposes. This study addresses image-to-image translation by convolutional neural networks (CNNs) to convert CBCT to CT-like scans, comparing supervised to unsupervised training techniques, exploiting a pelvic CT/CBCT publicly available dataset. Interestingly, quantitative results were in favor of supervised against unsupervised approach showing improvements in the HU accuracy (62% vs. 50%), structural similarity index (2.5% vs. 1.1%) and peak signal-to-noise ratio (15% vs. 8%). Qualitative results conversely showcased higher anatomical artifacts in the synthetic CBCT generated by the supervised techniques. This was motivated by the higher sensitivity of the supervised training technique to the pixel-wise correspondence contained in the loss function. The unsupervised technique does not require correspondence and mitigates this drawback as it combines adversarial, cycle consistency, and identity loss functions. Overall, two main impacts qualify the paper: (a) the feasibility of CNN to generate accurate synthetic CT from CBCT images, which is fast and easy to use compared to traditional techniques applied in clinics; (b) the proposal of guidelines to drive the selection of the better training technique, which can be shifted to more general image-to-image translation.

Keywords: image-to-image translation; synthetic images; supervised training; unsupervised training; U-Net; cycleGAN; CBCT; CT



Citation: Rossi, M.; Cerveri, P. Comparison of Supervised and Unsupervised Approaches for the Generation of Synthetic CT from Cone-Beam CT. *Diagnostics* **2021**, *11*, 1435. <https://doi.org/10.3390/diagnostics11081435>

Academic Editor: Sameer Antani

Received: 24 June 2021

Accepted: 7 August 2021

Published: 9 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Over the past few years, evidence has been accumulating in support of the view that machine learning and artificial intelligence methods may sensibly impact diagnostic and therapeutic areas. In this regard, clinical tools have been recently proposed in oncology to speeding up lesion analysis, improving tumor staging, supporting treatment planning, and, ultimately, contributing to the main clinical decisions [1,2]. Such tools may incorporate deep neural networks for image analysis, within a complete end-to-end clinical pipeline, providing automatic image segmentation, image synthesis, and image-to-image translation [3–5]. Especially, in radiation therapy, image-to-image translation has attracted clinical interest for the potential of using cone-beam computed tomography (CBCT), instead of the conventional computed tomography (CT), for both diagnostics and planning purposes [6–9]. CBCT is traditionally used for imaging the patient in-room and registering intra- and inter-fractional changes in the target anatomy along the treatment course [1]. Whether the actual morphological changes significantly alter the anatomy on which the dosimetric distribution computed at the planning stage was based, a new CT scan is acquired to perform a planning update. Despite allowing for much faster imaging than CT fan-beam imaging, the CBCT technique introduces a significant amount of scattered radiations, resulting in image artifacts such as shading, cupping, reduced image contrast, and beam-hardening [10,11]. Moreover, the pixel values in CBCT images may fluctuate because of these artifacts, making them not directly usable for dose calculation unless some correction methods are applied to calibrate CBCT to CT scanner Hounsfield Unit

(HU) values [12–14]. Thus, the availability of reliable techniques for the correction of CBCT images would extend their clinical usage, such as tumor shrinkage and organ shift evaluation, and treatment planning purposes, such as soft tissue segmentation for dose delivery computation. In addition, CBCT modality is compatible from a patient safety point of view (less additional non-therapeutic dose compared to traditional CT) with fractional dose delivery [1]. In this paper, generative deep neural networks were investigated for the translation of CBCT volumes into synthetic CT featuring compatible HU values. The same generative architecture was trained using supervised and unsupervised training techniques, exploiting a pelvic CT/CBCT publicly available dataset [15]. The comparison between the two obtained solutions was first quantified in terms of synthetic CBCT to original CT similarity by means of peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM). The HU agreement was then evaluated in terms of mean absolute error (MAE).

1.1. Background

Correction and translation methods for quantitative CBCT can be broadly divided into hardware-based, prior-CT-based, and model-based approaches. Hardware-based methods remove scatter based on additional equipment mounted on the acquisition system. In particular, an anti-scatter grid and partial beam blockers, along with software estimation of scattering, was proposed for scatter suppression [16–20]. While this kind of approach has shown promising results, it is not always feasible to be implemented in clinical practice due to the installation and setup of additional devices on acquisition systems. Moreover, this type of solution inherently reduces the system quantum efficiency and can degrade image quality [21]. Prior-CT-based methods leverage information obtained from high-resolution planning CT applied to up-to-date information contained in CBCT. They can be based on processing techniques like deformable registration of CT to CBCT [22], or by linearly scaling the CBCT HU values to CT ones through histogram matching [23,24]. The shading artifacts inside CBCT can also be estimated by low-pass filtering the difference between the high-resolution CT projections and the raw CBCT ones and then subtracted from the CBCT during reconstruction [25]. In general, these methods achieved a limited solution to CBCT problems since they depend on the accuracy of the spatial alignment between CBCT and planning CT volume pairs from the same patients. This alignment is not always possible, particularly in the case of acquisition made on different days or in the presence of soft tissues. Model-based methods typically use Monte Carlo simulations to model the scatter contribution to CBCT projection data. Many researchers attempt to model the scatter distribution via some combination of analytical methods or Monte Carlo algorithms [26–29]. Model-based methods demonstrated to reproduce HU to sufficiently robust accuracy for clinical application because they rigorously simulate photon transport. However, their primary limitation comes from the physics models. Moreover, these techniques are limited by their execution time, usually in the order of minutes to hours, making them incompatible with an online application (e.g., dose evaluation or pre-treatment adaptation). In recent years, deep learning-based methods joined traditional approaches to improve the quality of medical images. These methods leverage the advance in the field of image-to-image translation, making use of convolutional neural networks (CNNs) to learn how to map voxels from a source distribution (e.g., CBCT) to a target one (e.g., planning CT). Rather than truly correcting scattering or other physical effects, CNN learns this type of nonlinear mapping iteratively. Therefore, this approach does not need complex physical model formulation required by, e.g., Monte Carlo-based methods. Moreover, although deep learning-based methods can be computationally expensive to train, once a well-trained model is developed, the image correction can be applied in seconds, making these methods compatible with the clinical practice. The literature presents two main paradigms to train CNN for image-to-image translation: supervised and unsupervised training. Supervised training requires paired images from two dominions (e.g., CBCT and CT) for model training. The input CBCT is processed by the model generating a synthetic

CT (sCT), which is then compared to the corresponding ground-truth CT to minimize their pixel-by-pixel difference iteratively. Thanks to its neural encoding-decoding capabilities, the U-Net architecture is the most common one for this type of training [30–33]. Similar to prior-CT-based methods, images must be well-registered to achieve good performance when using supervised training. Since model training is based on a single pixel-wise loss function, the network could be biased and will tend to learn incorrect mapping in case of bad registration. It is not always possible to achieve such a good accuracy, especially in soft tissues, due to the patient's weight loss, tumor changes, or the presence of air. On the other hand, unsupervised training enabled the possibility to use unpaired data for image-to-image translation [34]. The most common architecture for this kind of training is cycle Generative Adversarial Network (cGAN) [35]. This architecture is based on two concurrent subnetworks, a generator and a discriminator, which work in opposition. Given two different datasets, the generator tries to learn the mapping to convert one dataset to the other. The generator aims to trick the discriminator. On the other hand, the discriminator objective is to distinguish between real and synthetic images. This generator-discriminator cycle-consistent loop is designed to improve the generator ability to produce synthetic images that are virtually indistinguishable from real ones. Cycle GAN was first proposed for natural image synthesis. Still, recently various researchers demonstrated its application for many medical image synthesis tasks, like the generation of synthetic CT from CBCT [36–39], MR synthesis from CT images [40,41] or PET attenuation correction [42]. The main advantage of this method consists of the possibility to use unpaired, even unbalanced datasets, as the one-to-one correspondence between both dominions is no longer necessary. However, the computational power needed to train this architecture increases dramatically, since there are four models to be trained based on at least four distinct loss functions.

1.2. Key Contributions of the Work

In summary, the traditional approaches presented some limitations. Hardware-based methods require the installation and the setup of additional devices on acquisition systems. These devices involve additional costs and do not always guarantee high flexibility and generalization capability. For what concerns prior-CT-based methods, their most significant limitation is the dependence on planning acquisitions and the accuracy of spatial alignment with CBCT. This condition is not always easily obtainable, especially when the acquisition period between the two modalities increase. Monte-Carlo-based methods are the most accurate approaches, but they suffer from two main limitations. Their execution time is usually incompatible with in-room clinical application, and they have the critical requirements of a complex physical model formulation. On the other hand, the deep-learning-based methods proposed in this work aim to overcome some of these limitations. At first, it does not have any additional hardware requirements, making it of potential utility for every kind of CBCT acquisition system, e.g., robotic C-arm, gantry mounted systems or couch-mounted ones. Moreover, while training a deep convolutional neural network is a time-demanding task, they are very fast in generating output images in a production application once trained. The execution time for these methods is compatible with clinical use. Furthermore, the black-box nature of neural networks does not rely in general on any complex physical modeling. In order to obtain a network with sufficient generalization capability, it is mandatory to train it with a sufficiently broad and generalized dataset. The first generative model was based on the U-Net architecture trained according to a traditional supervised pattern. The second generative model was still a U-Net-based model, which was coupled to a discriminator network and trained without supervision in a cGAN configuration. Overall, two main impacts qualify the paper: (a) the feasibility of CNN to generate accurate synthetic CT scans from CBCT images, which is fast and easy to use compared to traditional techniques applied in clinics; (b) the proposal of guidelines to drive the selection of the better training technique, which can be shifted to more general image-to-image translation.

2. Materials and Methods

2.1. Dataset Description

The publicly available dataset, called Pelvic Reference Dataset (PRD) [15], used in this work was obtained from the Cancer Imaging Archive (<https://www.cancerimagingarchive.net> accessed on 10 February 2021). PRD included paired CT and CBCT volumes of the pelvic region of 58 subjects, spanning 28 males and 30 females. CBCT images were acquired at least 1 week after the corresponding CT volume acquisition. Moreover, PRD includes 47 and 11 volumes acquired in supine and prone positions, respectively. Two subjects (both female, one prone, and one supine) were removed from the dataset. In the first case, a metal hip prosthesis made it impossible to distinguish the anatomical structures in both CT and CBCT. In the second case, the CBCT field of view was too small to contain the entire subject's pelvic region. Therefore, a massive part of the subject present in the CT was missing in its corresponding CBCT. These two cases were considered outliers and removed from the dataset, so that 56 cases were available for this study. Both CT and CBCT volumes featured size of 512×512 pixel on the axial plane with a pixel size of 1.00×1.00 mm and a single slice thickness of 3.00 mm. The number of slices of each CBCT was 88, while in CT scans that number was variable. Along with volume files, the PRD dataset also provided a table with the (x, y, z) coordinates to apply to the CBCT isocenter to align with their corresponding CT. To sum up, 4053 CT/CBCT 2D axial projections pairs were available. The entire dataset was split into train, validation, and test sets with a ratio of 80, 10, 10%, respectively, corresponding to 3243, 405, and 405 images per set.

2.2. Image Pre-Processing

Pre-processing steps were performed before feeding data to the neural network. At first, a binary mask was created to separate the subject from any non-anatomical content (e.g., treatment couch). This masking procedure avoids any negative impact from these structures on training procedures. In order to prepare these masks, Otsu thresholding was applied to each volume. A max connected component analysis was then performed on each mask to remove any non-anatomical residual finely. A final binary fill holes filtering and erosion filtering were applied to enhance mask accuracy further. Finally, the mask was applied to the original volume, obtaining an isolated anatomical region. Corresponding masked CT and CBCT were then rigidly registered, using isocenter coordinates provided with PRD Dataset. This step was necessary only for supervised training since the unsupervised one did not need paired data. However, good alignment was also required for evaluation analysis. The HU range of the grayscale values was first clipped to $[-1024, 3200]$ and then rescaled to $[0, 1]$ with a linear mapping. In order to reduce the computational cost for CNN, every axial slice was resampled to 256×256 pixel.

2.3. Deep Convolutional Neural Network Models

The two generative models were both based on the U-Net architecture trained according to two different training patterns. The supervised generative model required only this network, while the unsupervised one also required the definition of a discriminative architecture. In order to make the comparison possible, the generator architecture was precisely the same for both supervised and unsupervised procedures. The generator was implemented as an adapted version of the popular U-Net architecture. The basic U-Net is primarily used to solve pixel-by-pixel classification problems in image segmentation [43]. In the present work, U-Net was adapted to solve an image-to-image conversion task to generate sCT images from CBCT. The basic building blocks for the generator were depicted in Figure 1. The first one, called the ConvBlock, was based on a modified version of the implementation of Isola et al. [34]. It was composed of a 2D convolution with a 3×3 kernel, followed by an instance normalization layer and a swish activation function. Instance normalization was demonstrated to improve the performance in image generation tasks [44]. ConvBlock was present in the network both standalone and as a part of another processing block, called the InceptionBlock. This latter block was adapted from GoogLeNet [45]. It

comprised four parallel ConvBlocks, each with an increasing kernel size of dimensions 1×1 , 5×5 , 7×7 , and 11×11 . In this way, the input received by the InceptionBlock was processed simultaneously by multiple receptive fields. The output of each branch was then concatenated, and the entire stack of feature maps was returned as output. The main purpose of this processing block was to perform multi-scale feature extraction from the input image. These multi-scale extracted features, ranging from small to large receptive fields, can provide better results for image synthesis.

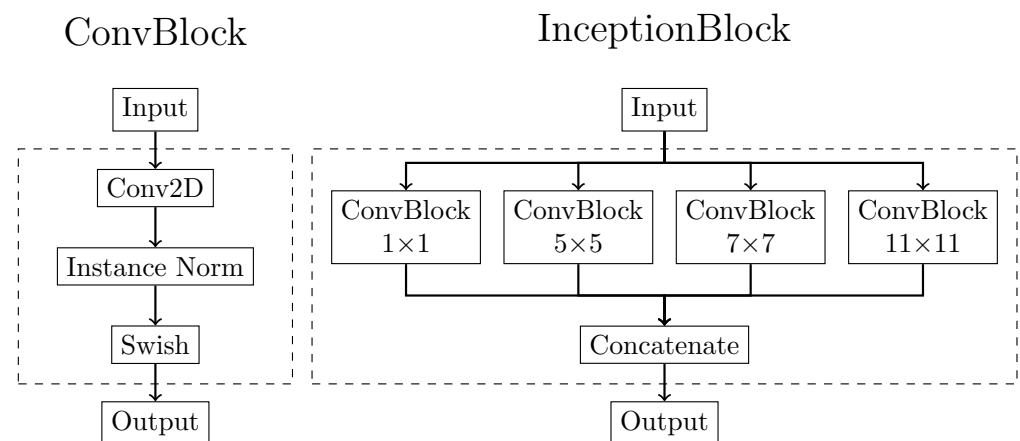


Figure 1. Basic building block architectures. The ConvBlock (left) is composed of a 2D convolutional step with kernel variable kernel size, followed by instance normalization and a Swish activation function. The InceptionBlock (right) is composed of the parallel combination of more ConvBlock with kernel dimension of 1×1 , 5×5 , 7×7 , and 11×11 . The output of each ConvBlock is then concatenated in a single output tensor.

The overall generator structure, depicted in Figure 2, was composed of a contracting and an expanding path, both based on the basic processing blocks. The first two upper processing blocks of the generator were composed of InceptionBlocks, while the deeper three were composed of ConvBlocks. In this way, the network was divided into two parts with two different functions: the inception part extracted global contextual information while the traditional part has the task to capture finer context and precise localization.

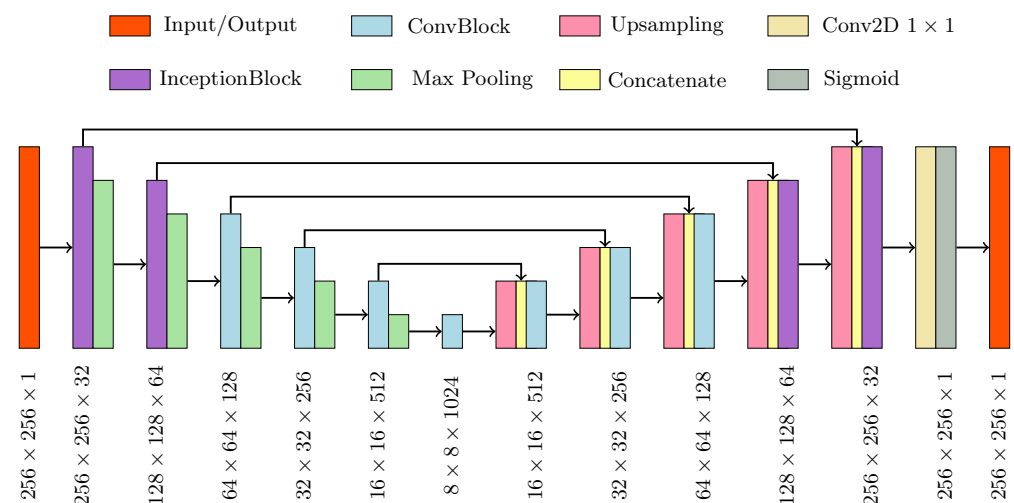


Figure 2. Schematic of the generator network architecture.

The discriminator (Figure 3) was a CNN that performs image classification. Its architecture was based on the PatchGAN architecture [34], considered to be the gold standard discriminator for cGAN [35,38,39]. It consisted of four consecutive ConvBlocks,

with a 4×4 kernel size. The convolution for the first three ConvBlock was set with stride 2, giving as outputs a tensor with half the size and twice the features map. The last ConvBlock had stride one and maintained the size and the number of feature maps unchanged. A sigmoid activation function followed the last layer, producing a 32×32 map with every pixel in the $[0, 1]$ range. This output map was used for patch-wise classification of the input image as real or fake.

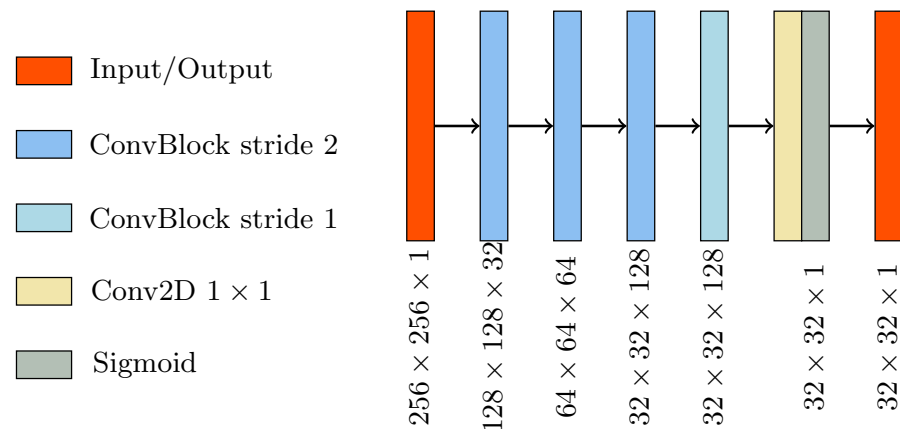


Figure 3. Schematic of the discriminator network architecture.

2.4. Training Methods

The two training routines are schematized in Figure 4. For supervised training, given the original CBCT image as input to generator CT (G_{CT}) network, the generated sCT is compared to the corresponding ground-truth CT. During training, the network adapts its weight according to MAE loss function, expressed as:

$$\mathcal{L}_{supervised}(CT, sCT) = \frac{1}{N} \sum_{i=0}^N |CT_i - sCT_i| \quad (1)$$

In regard to the unsupervised learning, the cGAN structure incorporated two generators and two discriminators, competing against one another, namely generator CT (G_{CT}), generator CBCT (G_{CBCT}), discriminator CT (D_{CT}), and discriminator CBCT (D_{CBCT}). G_{CT} was used to generate sCT from CBCT, while G_{CBCT} was used to generate synthetic CBCT (sCBCT) from CT. On the other hand, D_{CT} was used to distinguish between real CT and sCT, and D_{CBCT} is used to distinguish real CBCT from sCBCT. In the first step of the training, G_{CT} (G_{CBCT}) took CBCT (CT) as input and generated sCT (sCBCT). Then, G_{CT} (G_{CBCT}) took sCT (sCBCT) as input and generated a cycleCBCT (cycleCT), which is supposed to be equal to the original CBCT (CT). Meantime, D_{CT} (D_{CBCT}) tried to discriminate between real CT (CBCT), labeled as 1, and sCT (sCBCT), labeled as 0. Generator loss functions included three types of terms: adversarial loss, cycle consistency loss, and identity loss. Discriminator loss was composed only of an adversarial term. These loss functions were combined to mapping the distribution of the generated images to the distribution domain of the target images (see Appendix A).

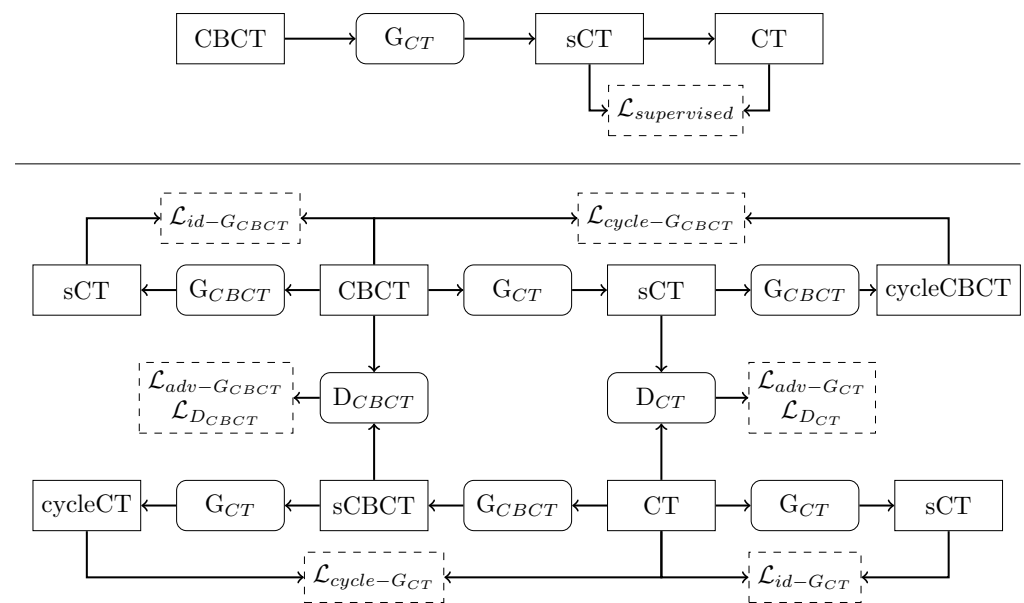


Figure 4. Schematic flow chart of supervised (**upper panel**) and unsupervised (**bottom panel**) training routines. Rectangle boxes represent images, and rounded corner boxes depict neural network models. Dashed boxes indicate a loss function. For a detailed explanation of the training routines, refer to the Appendix A.

2.5. Performance Metrics

In order to quantitatively evaluate network performances, three widely accepted numerical metrics were used: peak signal-to-noise ratio (PSNR), structural similarity index measure (SSIM), and mean absolute error (MAE). PSNR is computed as the ratio between the maximum possible power of a signal and the mean square error of the images being compared. Its value is measured in decibel and approaches infinity as the mean squared error between sCT and ground-truth CT approaches zero. Therefore, a higher PSNR value corresponded to higher image quality and vice-versa [46]. SSIM was first introduced in 2004 by Wang et al. [47]. It measures similarity between two images based on three factors: luminance, contrast, and structure. Compared to PSNR, SSIM provides an indication of similarity closest to that of the human visual system. The value for this metric ranges between 0 and 1, where 1 indicates the best possible level. Lastly, MAE is used to evaluate the HU accuracy between two images quantitatively. Before computing its value, the amplitude of the network output, i.e., [0,1], was scaled back to the original range, i.e., [−1024, 3200]. The lower the MAE, the higher the HU accuracy of the two images. Every metrics was evaluated using CT as the ground truth reference.

2.6. Cross-Validation Analysis

In order to validate the performance of the trained models, a four-fold cross-validation experiment was carried out on both methodologies. The entire dataset was divided into four subsets, each subgroup consisting of images of 14 patients. For each experiment, three subsets were used as the training set, while the remaining one was used as the test set. The cross-validation allowed a comparison of the performance of the models against the baseline results, i.e., the difference between the CT and the original CBCT images. The improvement in terms of SSIM, PSNR, and MAE was quantitatively analyzed. The statistical difference between the candidate models was also evaluated with Kruskal–Wallis non-parametric test for median differences ($p < 0.01$) and Tukey–Kramer post-hoc comparison.

2.7. Implementation Details

The Keras [48] and TensorFlow [49] Python frameworks were used to develop the network models, loss functions, metrics, and training routines. The training was carried out in the Google Colaboratory Cuda-enabled environment, equipped with a four-core CPU, 25 GB RAM, and NVIDIA® Tesla® P100 GPU support 16 GB RAM. The number of epochs has been set to 25. When the validation set SSIM score is maximized, the training routine was configured to save the set of the best network weights. The training was optimized with ADAM (Adaptive Moment Estimation) optimizer [50], with the following parameters: learning rate 2×10^{-4} , exponential decay rate for the first moment estimates $\beta_1 = 0.5$, and exponential decay rate for the second-moment estimates $\beta_2 = 0.999$. The batch size was set at 10. In order to generalize network performance as much as possible and prevent overfitting, data augmentation was performed at run-time during training. At the beginning of each training epoch, just before giving the image as input to the network, a series of image elaborations were applied. In particular, the images were randomly rotated by multiple of 90 degrees and horizontally flipped. The same transformations were also applied to the corresponding ground truth CT in order to maintain coherence between pairs. By applying these transformations, each CBCT/CT pair featured eight different configurations (four rotations times two flips). This implied that at each iteration the network was fed with different images. As an example, assuming that we have N images in the training dataset, at each iteration step of the training N different images are generated from the original samples applying the described random transformations .

3. Results

Supervised training required about 3 minutes per epoch. The only architecture used for this training pattern was the generator, corresponding to 2,554,977 trainable parameters. On the other hand, unsupervised training required about 5 minutes per epoch due to its increased complexity. This method featured two identical generators and two identical discriminators. Being the discriminator network characterized by 429,185 trainable parameters, the overall number of weights for the unsupervised model was 5,968,324. For what concern inference time, the two approaches were comparable. Generating sCT required less than 4 seconds for an entire CBCT volume (~70 slices) in both cases, computed on the same GPU environment used for training.

3.1. Performance Metrics

The cross-validation experiments confirmed that the proposed CNN models improved the metrics with respect to the baseline (Figure 5), being median and interquartile results summarized in Table 1. In general, supervised training attained better results for every metric. Considering SSIM, supervised sCT obtained an improvement of 2.5% with respect to the baseline ($p < 0.0001$), while unsupervised sCT gained an improvement of 1.1%. PSNR also confirmed this trend, resulting in an enhancement from the baseline for supervised sCT and unsupervised sCT of 4.19 dB ($p < 0.0001$) and 2.3 dB ($p < 0.0001$), respectively. The relative gains for PSNR were 15.7% and 8.6%. Lastly, the supervised model reduced the mean absolute error between supervised sCT and CT by 58.16 HU (62.3%), while the unsupervised model resulted in a reduction of 46.92 HU (50.3%). Even in this case, the supervised model appeared significantly better than the unsupervised one ($p < 0.0001$).

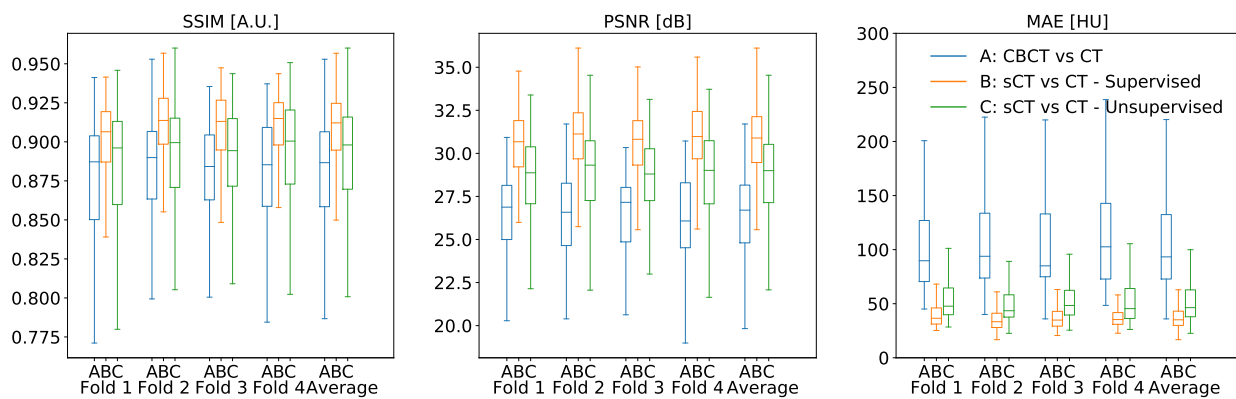


Figure 5. Quantitative analysis of SSIM, PSNR, and SSIM values between original CBCT, sCT supervised, and sCT unsupervised against the corresponding CT, computed for each fold of the four-fold cross-validation.

Table 1. Performance metrics evaluated on original CBCT, sCT supervised, and sCT unsupervised. Every value is computed against the original CT, considered as the ground truth. Every value is expressed as median (interquartile range).

	SSIM [A.U.]	PSNR [dB]	MAE [HU]
CBCT	0.887 (0.048)	26.70 (3.36)	93.30 (59.60)
Supervised sCT	0.912 (0.030)	30.89 (2.66)	35.14 (13.19)
Unsupervised sCT	0.898 (0.046)	29.00 (3.38)	46.38 (24.86)

3.2. Qualitative Comparison

Some examples of sCT generated by the supervised and unsupervised approaches are depicted in Figure 6. The first column represents the ground truth CT image, while the second one contains the original CBCT input. The third and fourth columns show the generated sCT images predicted from supervised and unsupervised models, respectively. The last column compares one intensity profile (the central row depicted with a line) for every imaging modality. The same example slices are also showed as difference maps in Figure 7. Every column represents the difference between the modality under evaluation (CBCT, supervised sCT, and unsupervised sCT) and the ground truth CT.

The first row represents the case of a CBCT/CT pair with a similar field of view (FOV). From the intensity profile and the difference map, it can be noticed that both models enhanced overall image qualities in terms of HU mapping, with slightly better visual details for the supervised model. The second row shows an example in which CT FOV is wider than CBCT one. In this case, it can be observed that the supervised model tried to compensate for the missing structures at the border with unrealistic values. Contours of supervised sCT were blurred and unreliable, while unsupervised sCT coped better with the original CBCT boundaries. In more detail, the unsupervised model did not attempt to recreate the missing FOV, resulting in more realistic and reliable contours with respect to the original CBCT. Concerning the internal anatomical structures, both models reached good qualitative results, as demonstrated by the corresponding intensity profile and difference map. The last row in Figures 6 and 7 shows a case in which the rectum area had an air-filled cavity with different shapes between the CT and the CBCT. The supervised model reacted by filling the air cavity with the surrounding values, while the unsupervised model preserved its contours. This latter behavior is preferable, as it is assumed that the information content of the CBCT is more updated than that of the CT. Again, the supervised model attempted to scale the HU values without considering possible changes in the anatomical structures of interest.

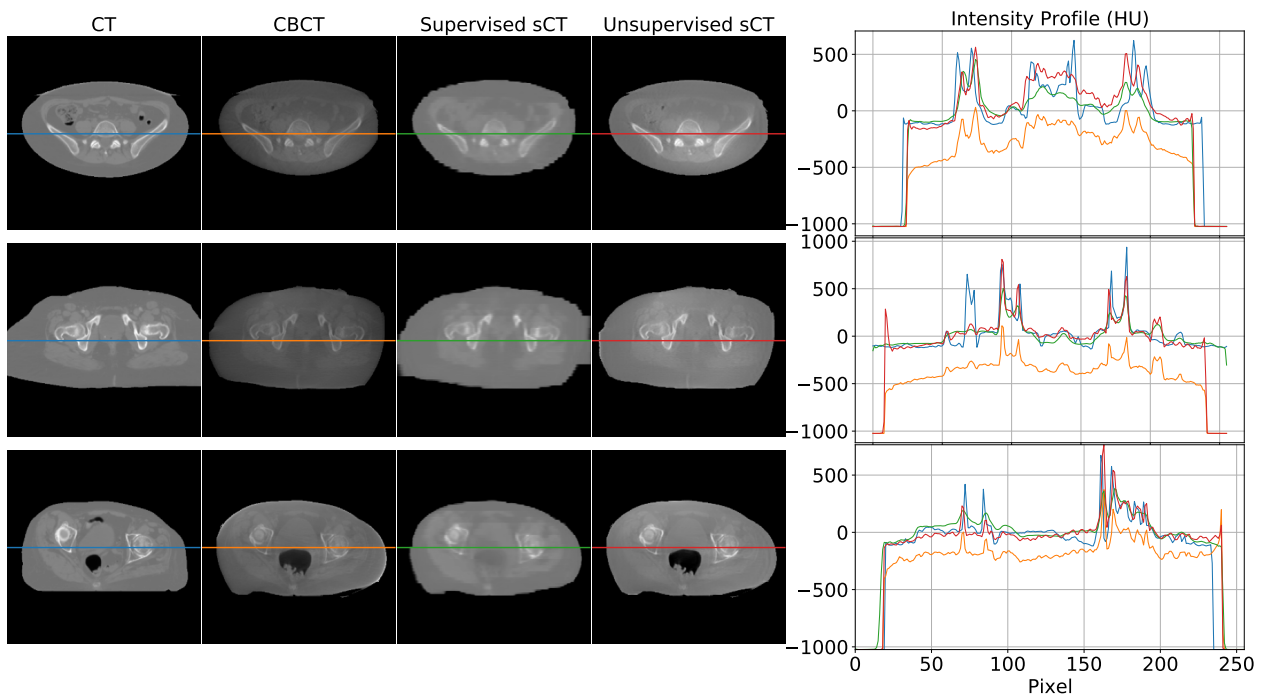


Figure 6. Visual comparison between some axial slices depicted in every modality: CT, CBCT, sCT supervised, and sCT unsupervised. Every row corresponds to a different example taken from a different subject. The rightmost part of the figure compares the intensity profiles of the central line of the images, highlighted by the central line in the four representations. Images are displayed with Window = 2000, Level = 0.

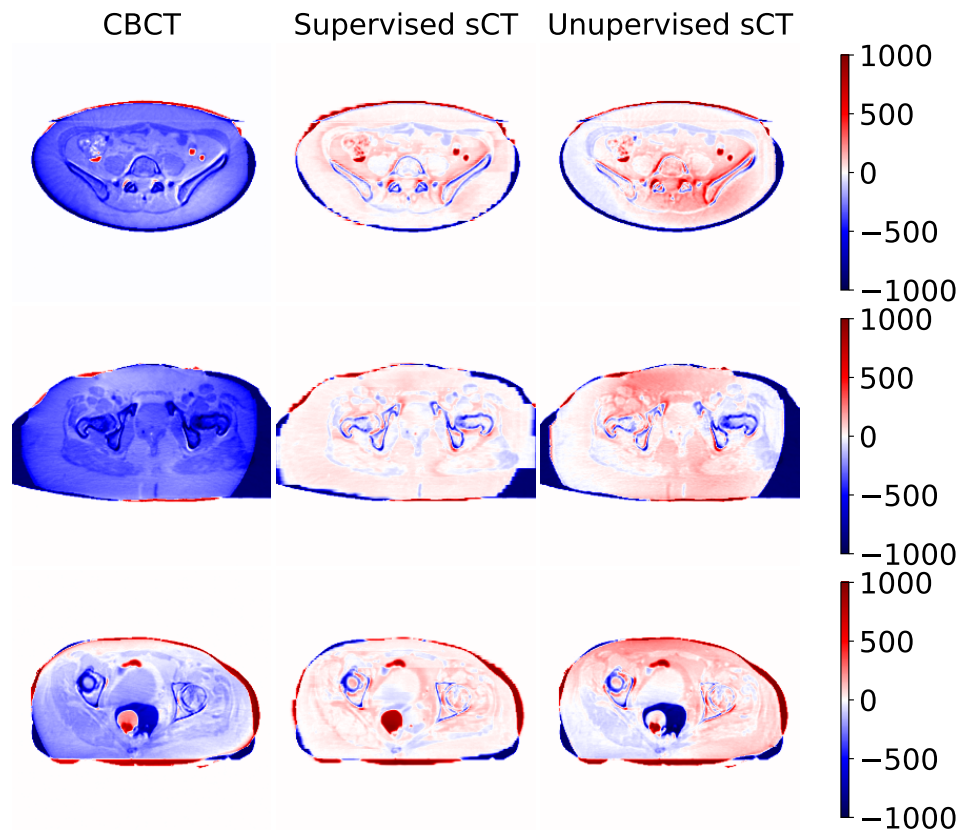


Figure 7. Difference maps computed between original CBCT, sCT supervised, and sCT unsupervised using their corresponding CT as reference. Numeric values are in Hounsfield Units. Each row represents a different example taken from a different subject. The examples depicted in this figure correspond to the example presented in Figure 6.

4. Discussion

4.1. Main Findings

This work proposed a deep-learning-based approach for generating synthetic CT images from CBCT scans, featuring HU values compatible with the traditional CT domain, providing a fair comparison between supervised and unsupervised training paradigms. As demonstrated by the cross-validation analysis, the supervised model outperformed the unsupervised ones. The supervised network guaranteed advantages in terms of computational cost, having 57% fewer weights to train and requiring 40% less training time for the same number of epochs. Moreover, it gained an increase of 12% in terms of MAE with respect to the unsupervised one, showing a superior ability in mapping HU values. On the other hand, the supervised model showed to be not always reliable, resulting in some unexpected behavior, especially in the contour regions. This kind of artifact did not appear in images generated through the unsupervised model, which proved to be more reliable in preserving the anatomical structure and the up-to-date information contained in the CBCT scan. One example of this is clearly visible in the bottom row of Figure 6, where the shape of the air cavity present in CBCT is preserved only for the unsupervised case, while the others tended to close the air gap. The unsupervised model did not suffer from this condition due to the cyclic loss functions, consisting of three main terms, each with a different purpose. The more straightforward loss function defined for the supervised model, i.e., mean square error, specialized the network to learn mapping without adequately considering the anatomical structure.

4.2. Comparison with the Literature

To our best knowledge, just one comparison between supervised and unsupervised training of deep networks, addressing back and forth translation between CT and magnetic resonance images was proposed in the recent literature [41]. In such a work, the authors explored different network architectures proposing a modified U-Net (supervised training) and the ResNet model for the generator (unsupervised training) proposed in [35]. Conversely, in our work, the same model was adopted in both the supervised and unsupervised generative parts, which smoothed out any biases in the comparison. Approaches competitive with our work were reported in the recent literature. Using an MC-based methodology, HU correction of about 31% for five lung cancer patient images was gained [13]. Another proposal, which used histogram matching on ten prostate cancer patient scans, resulted in a 20% HU correction [24]. A phantom-based study was carried out on the Catphan 600 (The Phantom Laboratory, Salem, NY, USA) [25] achieved an overall accuracy of 95% in HU recovery. Nonetheless, results on patient images were not presented. Concerning deep-learning-based methods, a proposal based on the U-Net trained with a supervised approach on 30 prostate cancer patients presented an HU accuracy enhancement of about 69% [30]. A comparative study explored the performances of some supervised models trained on 42 prostate cancer patients. The results obtained by the three tested models were 55, 17, 47% in terms of HU relative improvement [33]. Exploring the use of cGAN and unsupervised training, a study disclosed an enhancement of 57% with a dataset containing 120 head and neck scans [36]. Another unsupervised-based research resulted in a 16% increase for 33 prostate cancer patients [37]. The results proposed in this work are generally in line with the prior art, with an HU improvement of 62% and 50% for the supervised and unsupervised models, respectively.

4.3. Technical Challenges and Work Limitations

The use of complex models for remapping HU can be considered oversized when looking at the intensity profiles (cfr. Figure 6), where linear shifts appear predominant. This would suggest that regression models may be enough to address HU mapping. However, the inter-variability between the two modalities (e.g., air gaps, morphological variations due to postures, lesion progression) may easily introduce non-linearity in the intensity profile mapping (cfr. Figure 8). In the upper panel of Figure 8, an air gap can be appreci-

ated in the CBCT and in the corresponding CT. The notch in the intensity profiles is just produced by such a gap. This kind of deformation is difficult to correct by simple linear mapping. In the lower panel, relevant beam hardening of the CBCT at the image boundaries, corresponding to peripheral density appearance, is traceable in the corresponding intensity profile (orange curve) as an undue convexity at the borders. Likewise, a linear transform of the profile does not ensure accurate correction. Consequently, convolutional neural networks represent a natural approach to address as they feature input/output non-linear transforms and good generalization capabilities.

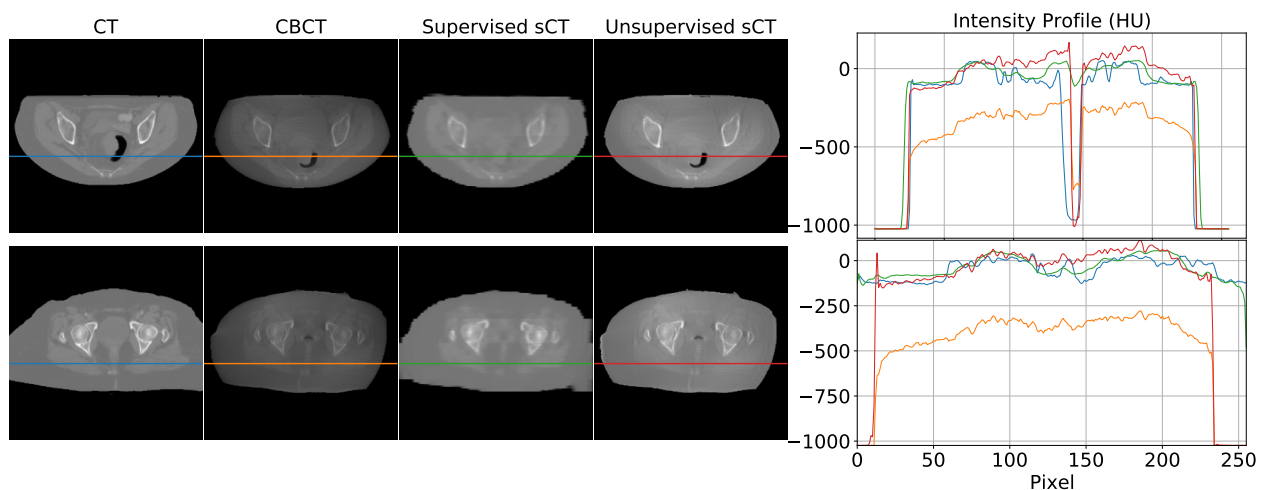


Figure 8. Examples of non-linearity of CBCT scans. Air gap introduced a notch in the intensity profile (**top**). Beam hardening is visible as a convex shape in the orange curve of both examples (**bottom**). Images are displayed with Window = 2000, Level = 0.

As we have shown, quantifying the image translation effectiveness by means of SSIM and PSNR can be inconclusive. For example, both lower SSIM and PSNR values (supervised against unsupervised) do not necessarily indicate that the images are qualitatively worse (cfr. Table 1). In particular, SSIM is a perception-based metric considered to be correlated with the quality perception of the human visual system, and the higher its value is, the better is the perception of similarity for human eyes [47]. Still, some studies have shown that SSIM is less effective when used to evaluate blurry and noisy images [51,52]. This reduces the relevance of comparisons performed using such metrics. Conversely, MAE provided an objective quantification of the HU matching. As far as the selection between the two approaches is concerned, we may argue that the supervised approach is preferable in the presence of a large amount of paired data, ensuring high accuracy in HU mapping. Nonetheless, gaining this advantage requires that CBCT and CT images be gathered as close as possible to minimize the anatomical difference. It is also crucial to mask images according to the smallest FOV between the two modalities in order to have comparable information content in between. In detail, masking ensures similar FOVs preventing the supervised network from attempting to reconstruct the sCT in the CBCT regions where this information is missing. In order to reduce the dependency on such constraints, the unsupervised method is suitable, avoiding pairing, requiring fewer pre-processing steps (e.g., masking) and being more conservative in terms of anatomical features. Removing the constraint on the pairing, the training effectiveness can benefit from a large amount of data, easier to collect. Nonetheless, even though obtaining unpaired images is clinically simpler, using explicitly paired data in training reduces geometric discrepancies at body boundaries. As verified in [21], unsupervised training may take advantage of paired data allowing to better focus on soft-tissue contrast than geometric mismatches. To further validate the methods presented in this work, a dosimetric analysis would be required. However, the chosen dataset did not allow this type of study to be explored due to the lack of the original CT dose distribution plan. Moreover, our methodologies focused only

on a single anatomic district, the pelvic one. We wanted to demonstrate the feasibility of both methods in a body region particularly subjected to inter- and intra-fractional changes. The comparison between the two training strategies was made to show their different approach in addressing image-to-image translation tasks, highlighting their strengths and limits.

5. Conclusions

We have shown that the proposed methodologies allow generating accurate synthetic CT images by automatically processing CBCT scans, featuring processing time faster than traditional Monte-Carlo-based methods. The basic idea behind this approach is to correct CBCT images exploiting their potential not only in a more accurate patient positioning, but also for clinical evaluation purposes as tumor shrinkage and organ shift. For what concerns the integration in the clinical workflow of treatment re-planning, this method can be integrated into a tool to be deployed that elaborates the reconstructed CBCT images, acting as an image filtering process. The correction of CBCT also makes possible the direct morphological comparison with the planning CT. As a consequence, the dose planning update, according to the new patient condition following fractional treatment, is just one of the possible usages of the CBCT, corrected with realistic tissue density according to HU. The clear advantage of using the in-room calibrated CBCT images for online re-planning consists of the availability of the actual patient anatomy. Nonetheless, this is still a challenge in radiation therapy and the subject of extensive research, which we plan to investigate in future works.

Author Contributions: Conceptualization, P.C.; data curation, M.R.; formal analysis, M.R.; funding acquisition, P.C.; methodology, M.R., P.C.; project administration, P.C.; resources P.C.; software programming, M.R.; supervision, P.C.; validation, M.R.; visualization preparation, M.R., P.C.; writing—original draft, M.R.; writing—review and editing, M.R., P.C. Both authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are openly available in The Cancer Imaging Archive (TCIA) at [10.7937/TCIA.2019.woskq500](https://www.cancerimagingarchive.net/), accessed on 23 June 2021. These collections are freely available to browse, download, and use for commercial, scientific and educational purposes as outlined in the Creative Commons Attribution 3.0 Unported License.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CBCT	Cone-Beam Computed Tomography
cGAN	cycle Generative Adversarial Network
CNN	Convolutional Neural Network
CT	Computed Tomography
G_{CT}	Generator CT
G_{CBCT}	Generator CBCT
D_{CT}	Discriminator CT
D_{CBCT}	Discriminator CBCT
FOV	Field of View
MAE	Mean Absolute Error
MSE	Mean Squared Error
PRD	Pelvic Reference Dataset
PSNR	Peak Signal-to-Noise Ratio
SSIM	Structural Similarity Index Measure

Appendix A

Loss Functions for Unsupervised Training

In the unsupervised training, the generator loss functions included three types of terms: adversarial loss, cycle consistency loss, and identity loss. Discriminator loss was composed only of an adversarial term. For G_{CT} the adversarial loss term is the mean squared error (MSE) between 1 and $D_{CT}(G_{CT}(CBCT))$:

$$\mathcal{L}_{adv-G_{CT}}(CBCT) = \frac{1}{N} \sum_{i=0}^N (1 - D_{CT}(G_{CT}(CBCT_i)))^2 \quad (A1)$$

Symmetrically, the adversarial objective for G_{CBCT} is:

$$\mathcal{L}_{adv-G_{CBCT}}(CT) = \frac{1}{N} \sum_{i=0}^N (1 - D_{CBCT}(G_{CBCT}(CT_i)))^2 \quad (A2)$$

The second term is represented by cycle consistency loss. It is used during training to enforce consistency of the two mappings from CT to sCBCT and CBCT to sCT [35]. This term is represented by the mean absolute error (MAE) between real CT (CBCT) and cycleCT (cycleCBCT), respectively:

$$\mathcal{L}_{cycle-G_{CT}}(CT) = \frac{1}{N} \sum_{i=0}^N |G_{CT}(G_{CBCT}(CT_i)) - CT_i| \quad (A3)$$

$$\mathcal{L}_{cycle-G_{CBCT}}(CBCT) = \frac{1}{N} \sum_{i=0}^N |G_{CBCT}(G_{CT}(CBCT_i)) - CBCT_i| \quad (A4)$$

The third term, identity mapping loss, was introduced in order to preserve HU values between real CT and sCT and between real CBCT and sCBCT. Without this term, generator G_{CT} and G_{CBCT} are free to change the HU values of input images even when it is not necessary. This loss acts as an identity constraint for the generator. When sCT is used as input for G_{CT} , it should generate CT too as output. The same applies for G_{CBCT} with CBCT. Therefore, identity loss is expressed as MAE between G_{CT} (CT) with respect to CT and between G_{CBCT} (CBCT) against real CBCT:

$$\mathcal{L}_{id-G_{CT}}(CT) = \frac{1}{N} \sum_{i=0}^N |G_{CT}(CT_i) - CT_i| \quad (A5)$$

$$\mathcal{L}_{id-G_{CBCT}}(CBCT) = \frac{1}{N} \sum_{i=0}^N |G_{CBCT}(CBCT_i) - CBCT_i| \quad (A6)$$

Combining all these terms, for the overall loss of the two generators the following formula holds:

$$\mathcal{L}_G = \mathcal{L}_{adv} + \lambda_{cycle} \cdot \mathcal{L}_{cycle} + \lambda_{id} \cdot \mathcal{L}_{id} \quad (A7)$$

where $\lambda_{cycle} = 10$ and $\lambda_{id} = 5$ represents the relative importance of the adversarial, cycle consistency, and identity mapping part of the objective.

The adversarial loss of the discriminator, which constrained the network to classify real images, labeled as 1, and synthetic images, labeled as 0, can be written as:

$$\mathcal{L}_{D_{CT}}(CT, CBCT) = \frac{1}{N} \sum_{i=0}^N \frac{(1 - D_{CT}(CT_i) + D_{CT}(G_{CT}(CBCT_i)))^2}{2} \quad (A8)$$

$$\mathcal{L}_{D_{CBCT}}(CT, CBCT) = \frac{1}{N} \sum_{i=0}^N \frac{(1 - D_{CBCT}(CBCT_i) + D_{CBCT}(G_{CBCT}(CT_i)))^2}{2} \quad (A9)$$

References

1. Ding, G.X.; Duggan, D.M.; Coffey, C.W.; Deeley, M.; Hallahan, D.E.; Cmelak, A.; Malcolm, A. A study on adaptive IMRT treatment planning using kV cone-beam CT. *Radiother. Oncol.* **2007**, *85*, 116–125. [[CrossRef](#)] [[PubMed](#)]
2. Niu, T.; Al-Basheer, A.; Zhu, L. Quantitative cone-beam CT imaging in radiation therapy using planning CT as a prior: First patient studies. *Med. Phys.* **2012**, *39*, 1991–2000. [[CrossRef](#)] [[PubMed](#)]
3. Lei, T.; Wang, R.; Wan, Y.; Du, X.; Meng, H.; Nandi, A.K. Medical image segmentation using deep learning: A survey. *arXiv* **2020**, arXiv:2009.13120.
4. Saba, T. Recent advancement in cancer detection using machine learning: systematic survey of decades, comparisons and challenges. *J. Infect. Public Health* **2020**, *13*, 1274–1289. [[CrossRef](#)]
5. Albertini, F.; Matter, M.; Nenoff, L.; Zhang, Y.; Lomax, A. Online daily adaptive proton therapy. *Br. J. Radiol.* **2020**, *93*, 20190594. [[CrossRef](#)]
6. Fattori, G.; Riboldi, M.; Pella, A.; Peroni, M.; Cerveri, P.; Desplanques, M.; Fontana, G.; Tagaste, B.; Valvo, F.; Orecchia, R.; Baroni, G. Image guided particle therapy in CNAO room 2: Implementation and clinical validation. *Phys. Med.* **2015**, *31*, 9–15. [[CrossRef](#)] [[PubMed](#)]
7. Veiga, C.; Janssens, G.; Teng, C.L.; Baudier, T.; Hotoiu, L.; McClelland, J.R.; Royle, G.; Lin, L.; Yin, L.; Metz, J.; et al. First clinical investigation of Cone Beam Computed Tomography and deformable registration for adaptive proton therapy for lung cancer. *Int. J. Radiat. Oncol. Biol. Phys.* **2016**, *95*, 549–559. [[CrossRef](#)]
8. Hua, C.; Yao, W.; Kidani, T.; Tomida, K.; Ozawa, S.; Nishimura, T.; Fujisawa, T.; Shinagawa, R.; Merchant, T.E. A robotic C-arm cone beam CT system for image-guided proton therapy: Design and performance. *Br. J. Radiol.* **2017**, *90*, 20170266. [[CrossRef](#)]
9. Landry, G.; Hua, C.h. Current state and future applications of radiological image guidance for particle therapy. *Med. Phys.* **2018**, *45*, e1086–e1095. [[CrossRef](#)]
10. Joseph, P.M.; Spital, R.D. The effects of scatter in x-ray computed tomography. *Med. Phys.* **1982**, *9*, 464–472. [[CrossRef](#)]
11. Schulze, R.; Heil, U.; Groß, D.; Bruellmann, D.D.; Dranischnikow, E.; Schwanecke, U.; Schoemer, E. Artefacts in CBCT: A review. *Dentomaxillofac. Radiol.* **2011**, *40*, 265–273. [[CrossRef](#)]
12. Kurz, C.; Kamp, F.; Park, Y.K.; Zöllner, C.; Rit, S.; Hansen, D.; Podesta, M.; Sharp, G.C.; Li, M.; Reiner, M.; et al. Investigating deformable image registration and scatter correction for CBCT-based dose calculation in adaptive IMPT. *Med. Phys.* **2016**, *43*, 5635–5646. [[CrossRef](#)]
13. Thing, R.S.; Bernchou, U.; Mainegra-Hing, E.; Hansen, O.; Brink, C. Hounsfield unit recovery in clinical cone beam CT images of the thorax acquired for image guided radiation therapy. *Phys. Med. Biol.* **2016**, *61*, 5781. [[CrossRef](#)]
14. Giacometti, V.; Hounsfield, A.R.; McGarry, C.K. A review of dose calculation approaches with cone beam CT in photon and proton therapy. *Phys. Med.* **2020**, *76*, 243–276. [[CrossRef](#)] [[PubMed](#)]
15. Yorke, A.A.; McDonald, G.C.; Solis, D.; Guerrero, T. *Pelvic Reference Data [Dataset]*; Atlassian Confluence Open Source Project License: Sydney, Australia, 2019. [[CrossRef](#)]
16. Siewerdsen, J.H.; Moseley, D.J.; Bakhtiar, B.; Richard, S.; Jaffray, D.A. The influence of antiscatter grids on soft-tissue detectability in cone-beam computed tomography with flat-panel detectors. *Med. Phys.* **2004**, *31*, 3506–3520. [[CrossRef](#)]
17. Zhu, L.; Xie, Y.; Wang, J.; Xing, L. Scatter correction for cone-beam CT in radiation therapy. *Med. Phys.* **2009**, *36*, 2258–2268. [[CrossRef](#)] [[PubMed](#)]
18. Sun, M.; Star-Lack, J.M. Improved scatter correction using adaptive scatter kernel superposition. *Phys. Med. Biol.* **2010**, *55*, 6695–6720. [[CrossRef](#)] [[PubMed](#)]
19. Sisniega, A.; Zbijewski, W.; Badal, A.; Kyprianou, I.S.; Stayman, J.W.; Vaquero, J.J.; Siewerdsen, J.H. Monte Carlo study of the effects of system geometry and antiscatter grids on cone-beam CT scatter distributions. *Med. Phys.* **2013**, *40*, 051915. [[CrossRef](#)]
20. Stankovic, U.; Ploeger, L.S.; van Herk, M.; Sonke, J.J. Optimal combination of anti-scatter grids and software correction for CBCT imaging. *Med. Phys.* **2017**, *44*, 4437–4451. [[CrossRef](#)] [[PubMed](#)]
21. Harms, J.; Lei, Y.; Wang, T.; Zhang, R.; Zhou, J.; Tang, X.; Curran, W.J.; Liu, T.; Yang, X. Paired cycle-GAN-based image correction for quantitative cone-beam computed tomography. *Med. Phys.* **2019**, *46*, 3998–4009. [[CrossRef](#)] [[PubMed](#)]
22. Zöllner, C.; Rit, S.; Kurz, C.; Vilches-Freixas, G.; Kamp, F.; Dedes, G.; Belka, C.; Parodi, K.; Landry, G. Decomposing a prior-CT-based cone-beam CT projection correction algorithm into scatter and beam hardening components. *Phys. Imaging Radiat. Oncol.* **2017**, *3*, 49–52. [[CrossRef](#)]
23. Abe, T.; Tateoka, K.; Saito, Y.; Nakazawa, T.; Yano, M.; Nakata, K.; Someya, M.; Hori, M.; Sakata, K. Method for converting Cone-Beam CT values into Hounsfield Units for radiation treatment planning. *Int. J. Med. Phys. Clin. Eng. Radiat. Oncol.* **2017**, *6*, 361–375. [[CrossRef](#)]
24. Kidar, H.S.; Azizi, H. Enhancement of Hounsfield unit distribution in cone-beam CT images for adaptive radiation therapy: Evaluation of a hybrid correction approach. *Phys. Med.* **2020**, *69*, 269–274. [[CrossRef](#)]
25. Niu, T.; Sun, M.; Star-Lack, J.; Gao, H.; Fan, Q.; Zhu, L. Shading correction for on-board cone-beam CT in radiation therapy using planning MDCT images. *Med. Phys.* **2010**, *37*, 5395–5406. [[CrossRef](#)] [[PubMed](#)]
26. Zbijewski, W.; Beekman, F.J. Efficient Monte Carlo based scatter artifact reduction in cone-beam micro-CT. *IEEE Trans. Med. Imaging* **2006**, *25*, 817–827. [[CrossRef](#)] [[PubMed](#)]
27. Bootsma, G.J.; Verhaegen, F.; Jaffray, D.A. Efficient scatter distribution estimation and correction in CBCT using concurrent Monte Carlo fitting. *Med. Phys.* **2014**, *42*, 54–68. [[CrossRef](#)]

28. Xu, Y.; Bai, T.; Yan, H.; Ouyang, L.; Pompos, A.; Wang, J.; Zhou, L.; Jiang, S.B.; Jia, X. A practical cone-beam CT scatter correction method with optimized Monte Carlo simulations for image-guided radiation therapy. *Phys. Med. Biol.* **2015**, *60*, 3567–3587. [[CrossRef](#)]
29. Zhao, W.; Vernekohl, D.; Zhu, J.; Wang, L.; Xing, L. A model-based scatter artifacts correction for cone beam CT. *Med. Phys.* **2016**, *43*, 1736–1753. [[CrossRef](#)] [[PubMed](#)]
30. Hansen, D.C.; Landry, G.; Kamp, F.; Li, M.; Belka, C.; Parodi, K.; Kurz, C. ScatterNet: A convolutional neural network for cone-beam CT intensity correction. *Med. Phys.* **2018**, *45*, 4916–4926. [[CrossRef](#)]
31. Maier, J.; Eulig, E.; Vöth, T.; Knaup, M.; Kuntz, J.; Sawall, S.; Kachelrieß, M. Real-time scatter estimation for medical CT using the deep scatter estimation: Method and robustness analysis with respect to different anatomies, dose levels, tube voltages, and data truncation. *Med. Phys.* **2019**, *46*, 238–249. [[CrossRef](#)]
32. Kida, S.; Nakamoto, T.; Nakano, M.; Nawa, K.; Haga, A.; Kotoku, J.; Yamashita, H.; Nakagawa, K. Cone Beam Computed Tomography image quality improvement using a deep convolutional neural network. *Cureus* **2018**, *10*, e2548. [[CrossRef](#)] [[PubMed](#)]
33. Landry, G.; Hansen, D.; Kamp, F.; Li, M.; Hoyle, B.; Weller, J.; Parodi, K.; Belka, C.; Kurz, C. Comparing Unet training with three different datasets to correct CBCT images for prostate radiotherapy dose calculations. *Phys. Med. Biol.* **2019**, *64*, 035011. [[CrossRef](#)] [[PubMed](#)]
34. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-To-Image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.
35. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2223–2232.
36. Liang, X.; Chen, L.; Nguyen, D.; Zhou, Z.; Gu, X.; Yang, M.; Wang, J.; Jiang, S. Generating synthesized computed tomography (CT) from cone-beam computed tomography (CBCT) using CycleGAN for adaptive radiation therapy. *Phys. Med. Biol.* **2019**, *64*, 125002. [[CrossRef](#)]
37. Kurz, C.; Maspero, M.; Savenije, M.H.F.; Landry, G.; Kamp, F.; Pinto, M.; Li, M.; Parodi, K.; Belka, C.; van den Berg, C.A.T. CBCT correction using a cycle-consistent generative adversarial network and unpaired training to enable photon and proton dose calculation. *Phys. Med. Biol.* **2019**, *64*, 225004. [[CrossRef](#)]
38. Kida, S.; Kaji, S.; Nawa, K.; Imae, T.; Nakamoto, T.; Ozaki, S.; Ohta, T.; Nozawa, Y.; Nakagawa, K. Visual enhancement of Cone-beam CT by use of CycleGAN. *Med. Phys.* **2020**, *47*, 998–1010. [[CrossRef](#)]
39. Tien, H.J.; Yang, H.C.; Shueng, P.W.; Chen, J.C. Cone-beam CT image quality improvement using cycle-deblur consistent adversarial networks (cycle-deblur GAN) for chest CT imaging in breast cancer patients. *Sci. Rep.* **2021**, *11*, 1133. [[CrossRef](#)]
40. Nie, D.; Trullo, R.; Lian, J.; Petitjean, C.; Ruan, S.; Wang, Q.; Shen, D. Medical image synthesis with context-aware generative adversarial networks. In Proceedings of the Medical Image Computing and Computer Assisted Intervention—MICCAI 2017, Quebec City, QC, Canada, 11–13 September 2017; pp. 417–425. [[CrossRef](#)]
41. Li, Y.; Li, W.; Xiong, J.; Xia, J.; Xie, Y. Comparison of supervised and unsupervised deep learning methods for medical image synthesis between Computed Tomography and Magnetic Resonance images. *BioMed Res. Int.* **2020**, *2020*, 5193707. [[CrossRef](#)]
42. Dong, X.; Lei, Y.; Wang, T.; Higgins, K.; Liu, T.; Curran, W.J.; Mao, H.; Nye, J.A.; Yang, X. Deep learning-based attenuation correction in the absence of structural information for whole-body positron emission tomography imaging. *Phys. Med. Biol.* **2020**, *65*, 055011. [[CrossRef](#)]
43. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing And Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.
44. Ulyanov, D.; Vedaldi, A.; Lempitsky, V. Instance normalization: The missing ingredient for fast stylization. *arXiv* **2016**, arXiv:1607.08022.
45. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1–9.
46. Hore, A.; Ziou, D. Image quality metrics: PSNR vs. SSIM. In Proceedings of the 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010. [[CrossRef](#)]
47. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [[CrossRef](#)]
48. Chollet, F. Keras. 2015. Available online: <https://keras.io> (accessed on 21 February 2021).
49. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, 2015. Available online: [tensorflow.org](https://www.tensorflow.org) (accessed on 21 February 2021).
50. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
51. hao Chen, G.; ling Yang, C.; li Xie, S. Gradient-based structural similarity for image quality assessment. In Proceedings of the 2006 International Conference on Image Processing, Atlanta, GA, USA, 8–11 October 2006. [[CrossRef](#)]
52. Li, C.; Bovik, A.C. Content-partitioned structural similarity index for image quality assessment. *Signal Process. Image Commun.* **2010**, *25*, 517–526. [[CrossRef](#)]