



**HAL**  
open science

## Indirect deformable image registration using synthetic image generated by unsupervised deep learning

Cedric Hemon, Blanche Texier, Hilda Chourak, Antoine Simon, Igor Bessieres, Renaud de Crevoisier, Joël Castelli, Caroline Lafond, Anais Barateau, Jean-Claude Nunes

### ► To cite this version:

Cedric Hemon, Blanche Texier, Hilda Chourak, Antoine Simon, Igor Bessieres, et al.. Indirect deformable image registration using synthetic image generated by unsupervised deep learning. *Image and Vision Computing*, 2024, 148, pp.105143. 10.1016/j.imavis.2024.105143 . hal-04651011

**HAL Id: hal-04651011**

**<https://hal.science/hal-04651011>**

Submitted on 17 Jul 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



# Indirect deformable image registration using synthetic image generated by unsupervised deep learning

Cédric Hémon<sup>a,\*</sup>, Blanche Texier<sup>a,1</sup>, Hilda Chourak<sup>a</sup>, Antoine Simon<sup>a</sup>, Igor Bessières<sup>b</sup>,  
Renaud de Crevoisier<sup>a</sup>, Joël Castelli<sup>a</sup>, Caroline Lafond<sup>a</sup>, Anaïs Barateau<sup>a</sup>, Jean-  
Claude Nunes<sup>a,\*</sup>

<sup>a</sup> Univ Rennes, CLCC Eugène Marquis, INSERM, LTSI - UMR 1099, F-35000 Rennes, France

<sup>b</sup> Centre Georges-François Leclerc, Dijon, France

## ARTICLE INFO

### Keywords:

Multimodal image registration  
Synthetic-CT  
MRI  
CBCT  
Radiotherapy  
Unsupervised generation

## ABSTRACT

**Background and purpose:** 3D image registration is now common in many medical domains. Multimodal registration implies the use of different imaging modalities, which results in lower accuracy compared to monomodal registration. The aim of this study was to propose a novel approach for deformable image registration (DIR) that incorporates an unsupervised deep learning (DL)-based generation step. The objective was to reduce the challenge of multimodal registration to monomodal registration.

**Material and methods:** Two datasets from prostate radiotherapy patients were used to evaluate the proposed method. The first dataset consisted of Computed Tomography (CT)/ Cone Beam Computed Tomography (CBCT) pairs from 23 patients using different CBCT devices. The second dataset included Magnetic Resonance Imaging (MRI)/CT pairs from two different care centers, utilizing different MRI devices (0.35 T MRIdian MR-Linac, 1.5 T GE lightspeed MRI). Following a preprocessing step essential for ensuring DL synthesis accuracy and standardizing the database, synthetic CTs ( $sCT_{reg}$ ) were generated using an unsupervised conditional Generative Adversarial Network (cGAN). The generated sCTs from CBCT or MRI were then utilized for deformable registration with CT scans. This registration method was compared to three standard methods: rigid registration, Elastix registration based on BSplines, and VoxelMorph-based registration (applied exclusively to CBCT/CT). The endpoints of comparison were the dice coefficients calculated between delineated structures for both datasets.

**Results:** For both datasets, intermediary sCT generation provided the highest dice coefficients. Dices reached 0.85, 0.85 and 0.75 for the prostate, bladder and rectum for the dataset 1 and 0.90, 0.95 and 0.87 respectively for the dataset 2. When the sCT were not used, dices reached 0.66, 0.78, 0.66 for the dataset 1 and 0.93, 0.87 and 0.84 for the dataset 2. Furthermore, the evaluation of the impact of registration on sCT generation showed that lower Mean Absolute Errors were obtained when the registration was conducted with a sCT.

**Conclusions:** Using unsupervised deep learning to synthesize intermediate sCT has led to improved registration accuracy in radiotherapy applications employing two distinct imaging modalities.

## 1. Introduction

3D deformable image registration has become common practice in various medical applications and imaging modalities. Radiotherapy, a technique employing high-energy radiation to target and eliminate cancer cells, often involves the acquisition of multiple images. Among these, a CT scan is typically obtained and designated as the reference image for dose calculation in formulating the treatment plan. To

accomplish this, a medical physicist utilizes specialized treatment planning software to design a radiation therapy plan. This process entails selecting the appropriate treatment beams, shaping them to conform to the tumor, and optimizing the dose distribution to maximize tumor control while minimizing side effects. Subsequently, as treatment delivery occurs over several days, a daily image such as Cone Beam Computed Tomography (CBCT) or Magnetic Resonance Imaging (MRI) [1] is acquired to register the target volume before irradiation.

\* Corresponding author.

E-mail address: [cedric.hemon@univ-rennes.fr](mailto:cedric.hemon@univ-rennes.fr) (C. Hémon).

<sup>1</sup> C. Hémon & B. Texier contributed equally to this work.

Nowadays, the required targeting accuracy is at the millimeter level. Image registration can occur at various stages of the radiotherapy workflow: 1) during treatment plan definition, using a reference CT and a secondary image (MRI, PET...), and 2) during each treatment session to align the patient/tumor position. The 3D daily image enables monitoring of inter-session variations and estimation of the delivered dose to the target volume (TV) and organs at risk (OARs) [2]. In the field of radiation therapy (RT), a significant challenge lies in achieving precise registration across different imaging modalities, known as multimodal registration, which often results in lower accuracy compared to monomodal registration. Additionally, given the variation in acquisition devices across centers (acquisition protocol, magnetic field, field of view), it is crucial to develop a training method that can be applied across multiple centers. Multicenter learning emerges as a promising solution to address this need [3].

Another recent interest in deformable image registration (DIR) in radiotherapy is the evaluation of the accuracy of generating synthetic CT ( $sCT_{eval}$ ) by comparing the intensities (measured in Hounsfield Units = HU) with those of a reference CT (deformed CT to correspond to the daily image). The generation of an  $sCT$  from the secondary image becomes essential for calculating the dose delivered to the patient [4]. In the context of MRI-only radiation therapy, the generation of  $sCT$ s by deep learning (DL) is widely proposed in the literature [4–6]. Most supervised DL-based synthesis methods [5] require accurate registration of the training image pairs to ensure optimal convergence. Moreover, for accurate voxel-to-voxel evaluation of the  $sCT_{eval}$ , precise registration is necessary. Any medical image registration is confronted with anatomical variations, which particularly depend on the acquisition delay between images. Moreover, multimodal registration presents additional challenges compared to monomodal registration due to differences in the intrinsic characteristics of imaging modalities and image quality (such as noise and artifacts) [7–10]. Different imaging techniques capture various aspects of anatomy and physiology, resulting in variations in image appearance, texture, and intensity. Artifacts can interfere with the accurate identification of corresponding features, leading to errors in the registration process [11].

The standard workflow of registration relies on four elements: the optimizer, the interpolation, the correlation measures, and the spatial transformer [12,13]. The most commonly used approach in multimodal registration involves the introduction of non-linear correlation measures [14,15]. These measures quantify the similarity between images by considering the non-linear relationships between their pixel values, thus enabling more robust registration of multimodal data. Despite their effectiveness, non-linear correlation measures have limitations. Their effectiveness may vary depending on the specific multimodal data types encountered, primarily due to the presence of many local maxima in these functions [16]. This prevalence of local maxima poses challenges for optimization methods, often resulting in registration errors.

In the field of DL-based registration, recent studies have led to innovative multimodal metrics [17]. These metrics leverage neural networks to estimate the similarity between the source and target images. The advent of these metrics based on neural networks has paved the way for the development of non-linear metrics, enabling registration processes that involve projecting the two images into a non-linear feature space [18], characterized by increased discriminating capability and reduced size. Despite their advantages, DL models come with several drawbacks, including their inherent complexity, lack of interpretability, and the necessity for substantial computational resources and extensive datasets.

To overcome these limitations and move away from relying on multimodal metrics, researchers have explored the concept of transforming multimodal problems into monomodal ones [19–23]. One approach involves generating an image of the target modality from the source image, which offers a potential solution to enhance model performance and interpretability. Several studies have suggested solving registration and synthesis tasks jointly using combined networks, as

highlighted in the works of Liu et al. [19], Han et al. [20], Deng et al. [21], and Zhou et al. [22]. In their study, Liu et al. performed MR/CT (pelvis and brain) registration steps prior to generation to minimize monomodal loss. Han et al. [20] undertook MR/CBCT registration (brain), by incorporating a generation step to align both modalities by projecting them into the CT definition space. Deng et al. [21] and Zhou et al. [22] employed a registration step to enhance the generation of  $sCT$  (head & neck) from MRI, aiming to mitigate registration faults during the learning process.

Using deep learning-based registration as their foundation, these approaches encounter the specific challenges outlined above. Additionally, they employ cycleGAN for unsupervised generation, which is a resource-intensive approach posing problems of instability and collapse modes [24]. In the natural image domain, while unsupervised contrastive learning and transfer learning have gained prominence, particularly as proposed in [25,26], contrastive learning remains less prevalent in medical image generation [27,28] in a monochrome context. Nevertheless, transfer learning, particularly for tasks such as segmentation, is becoming increasingly popular, helped by the emergence of pre-trained networks, such as TotalSegmentator [29]. Alternatively, Yang et al. [30] proposed using an intermediate  $sCT$  to achieve monomodal registration (CT/ $sCT$ ). To enable unsupervised training of their generation model, they performed data augmentations (combining spatial and intensity techniques) on their initially paired dataset (nine sets of brain MR-CT).

The main contributions and novelties of this paper are:

- An accurate unsupervised synthesis method based on a cGAN was used. This technique had previously been employed as part of the SynthRAD2023 challenge within a multicenter context (involving two different acquisition devices per modality) [31]. Indeed, the synthesis approach employed incorporates the CREPs loss function, using only the style term of the ConvNext-based perceptual loss.
- The multimodal registration problem is translated into a monomodal registration problem through a two-step process (generation, registration) in the pelvic area for two multimodal registration challenges: CBCT/CT and MRI/CT. However, to our knowledge, this has not yet been achieved. This technique is specifically applied to the pelvic region, which poses greater challenges compared to some other anatomical sites due to the dynamic movements induced by digestion and bladder filling.
- In contrast to the literature, the used model does not follow an end-to-end deep learning approach. Instead, it represents a combination of deep learning for generation and conventional methods for registration, overcoming the typical constraints associated with deep learning in registration tasks.
- An alternative solution is also explored for CBCT, aimed at eliminating the generation step. This alternative consists of using a deep learning network for registration estimation and employs a new multimodal and non-linear similarity measure between CBCT and CT. This metric is developed to be in line with the underlying principles of the proposed method, which integrates both content and style aspects.

## 2. Material and methods

In the field of radiation therapy, achieving multimodal DIR between MRI and CT images, as well as between CBCT and CT, is imperative. Commonly, multimodal registration methods rely on metrics such as mutual information, yet algorithms often encounter challenges in establishing spatial correspondences between these diverse modalities. In response to this issue and to improve the registration accuracy, this paper introduces an innovative approach: using image synthesis to translate the inherently multimodal issue into a more manageable monomodal one.

For the image synthesis process, supervised algorithms demonstrate greater accuracy. However, the efficiency of supervised deep learning

models is compromised due to the lack of registration in the training cohort. To address this limitation, recent studies propose unsupervised algorithms as a solution, aiming to obtain more accurate images despite the absence of explicit correspondences in the training set.

The Fig. 1 illustrates the registration method and the workflow of the study. The registration method was compared to rigid registration and other deformable registration methods. Two types of endpoints were used for the evaluation.

## 2.1. Dataset

Two datasets of prostate cancer patients were used in this study: one composed of 60 pairs of CT/MRI planning images, and the other composed of 23 patients, each with a planning CT and CBCTs acquired during sessions (daily or weekly acquisition), totaling 341 CBCTs.

Dataset 1 corresponds to the CT/MRI dataset, with images acquired at two different centers. The first 30 patients from Center 1 (C1) had CT scans acquired in a head-first supine position on a GE Light-SpeedRT16, and MRIs obtained on a 0.35 T MRIdian (ViewRay) MRI-Linac using a Trufi sequence. Center 2 (C2) comprised 30 patients with CT scans acquired in a head-first supine position on a Philips BigBore Scanner, and MRIs obtained on a 1.5 T Siemens Skyra MRI scanner, using a T2 sequence.

The dataset 2 corresponds to the DELPEL project [2]. This dataset includes 23 planning computed tomography (CT) and 341 cone beam CT (CBCT) images from 23 patients with localized prostate cancer. All planning CT scans were acquired in a head-first supine position with a rectal catheter and intravenous iodine contrast, covering the sacroiliac joints to the lower edge of the lesser trochanters, using a 3 mm slice thickness. The CBCT images were acquired in a head-first supine position either weekly (17 patients corresponding to 132 images) or daily (6

patients corresponding to 209 images) during the 8 weeks of treatment. The CBCT images were acquired using either Varian On-Board Imager ( $n = 16$ ) or Elekta XVI ( $n = 7$ ), with a 1 mm slice thickness. Patients were asked to have a full bladder at the time of treatment. All images were delineated by experts to obtain the prostate, bladder, and rectum structures.

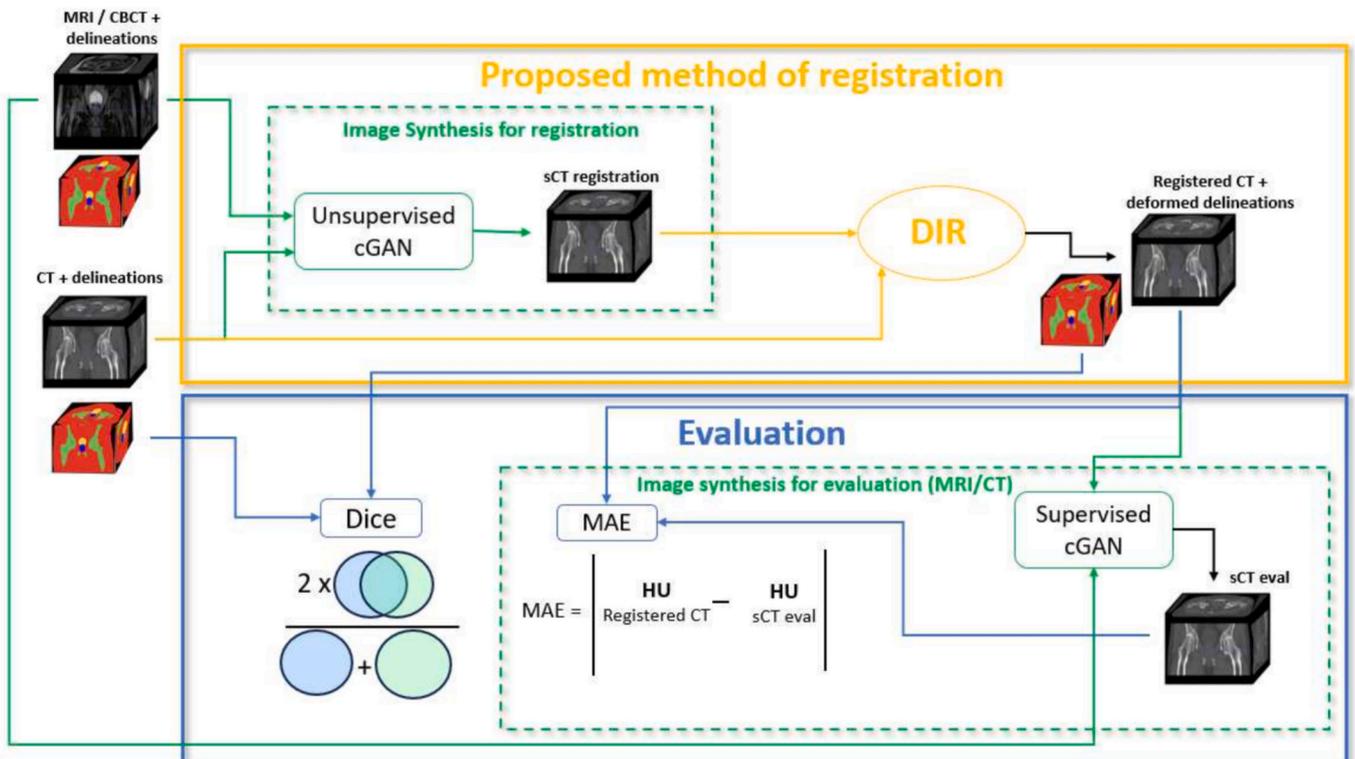
To quantify the need for deformation between modalities, the volumes of each structure were evaluated. Table 1 shows the intra-patient differences (with CT as the reference) in terms of volume for each structure between both imaging modalities.

## 2.2. Image preprocessing

For dataset 1, preprocessing steps were performed due to the heterogeneity between images. First, a threshold was applied to eliminate outlier values, with  $[-1000;1600]$  HU for the CTs and  $[0;500]$  for MRIs. Then, to correct MRI non-uniformity, images underwent preprocessing steps, including: (1) N4 bias field correction [32]; (2) histogram matching; and (3) filtering using gradient anisotropic diffusion [33]

**Table 1**  
Volumes differences of delineated structures (CT minus MRI/CBCT)  $\pm$  standard deviation.

Structures	Difference (in $cm^3$ )			Absolute difference (in $cm^3$ )		
	Prostate	Bladder	Rectum	Prostate	Bladder	Rectum
Dataset 1: MRI/CT	$7.9 \pm 23.6$	$99.7 \pm 217.0$	$1.6 \pm 31.3$	$14.4 \pm 20.5$	$159.0 \pm 177.4$	$19.8 \pm 24.2$
Dataset 2: CBCT/ CT	$5.9 \pm 9.3$	$-91.3 \pm 83.8$	$-14.17 \pm 42.1$	$6.7 \pm 8.7$	$96.6 \pm 83.8$	$29.5 \pm 33.2$



**Fig. 1.** Workflow of the proposed method for MRI or CBCT/CT registration. First, a learning dataset composed of preprocessed and rigidly registered MRI/CT or CBCT/CT pairs is used to synthesize a synthetic CT ( $sCT_{reg}$ ) with an unsupervised cGAN. Then, this  $sCT_{reg}$  is used to guide the deformable image registration (DIR) with the CT using the Elastix library based on BSplines. For the MRI dataset, the MRI and the registered CT are used to train a supervised 3D cGAN. Then, the different models of  $sCT_{eval}$  generation obtained with the different kinds of registration are evaluated with the mean absolute error (MAE) metric. Finally, the registrations (CBCT/CT and MRI/CT) were evaluated with dice coefficients on the manually delineated structures of body, bladder, prostate, and rectum.

(with 10 iterations, time step: 0.03, conductance: 1.0).

For dataset 2, to ensure proper patient positioning and to initiate the DIR process, the CBCTs were aligned with the CT anatomy through bone-based rigid registration. Subsequently, for each patient, both the CT and CBCT images were adjusted to match the smaller field of view of the CBCT and resized to a uniform size of 256x256x128 voxels. This resizing was accomplished using B-Spline interpolation, resulting in a consistent resolution for comparison.

Furthermore, this registration process was applied to a supervised DL-based sCT synthesis learning cohort. One of the main preprocessing steps for pelvic sCT synthesis is the standardization of field of views (FOV). A shared FOV between the two modalities improves coherence. FOV size was homogenized as follows: 12 cm above and below the prostate barycenter for the first  $sCT_{reg}$  generation before registration and 8 cm for the second generation to eliminate edge effects due to organ deformations from the registration. After each FOV unification and to achieve a uniform size of 256x256x128 voxels, B-Spline interpolation was used.

### 2.3. Unsupervised sCT generation to guide registration

The initial phase of registration of the two images from distinct modalities consisted of generation on sCT. This was achieved through unsupervised learning by using unpaired patient data and by not using a voxel-wise loss function on the ground truth (CT), thus eliminating the need for both a registration step and the use of a paired database.

#### 2.3.1. CBCT-to-CT synthesis

The proposed algorithm to generate  $sCT_{reg}$  from CBCT was an unsupervised 2D cGAN [34]. This approach used a customized 2D generator consisting of a 6-block residual network positioned between a single down-sampling block and an individual up-sampling block. No biases are incorporated into the convolution kernel throughout the network architecture. Following each convolutional operation, batch normalization is applied, subsequently followed by the activation function. A LeakyReLU activation with a fixed slope of 0.2 is employed for the ResNet-Block, while ReLU activation is used for other blocks, except for the final convolution, where no activation function is applied. Binary Cross Entropy is used as the discriminator loss (a PatchGAN).

To generate sCT in an unsupervised manner, avoiding the introduction of registration bias and to address the instability and collapse issues resulting from the absence of an explicit loss function [35], the sCT generation process has been approached as a style transfer problem. Neural style transfer, considered as a domain adaptation technique [36] for unsupervised training, reduces these challenges [37] through the incorporation of perceptual loss, for instance.

The unsupervised network was trained on CT and CBCT of two distinct patients. CBCT/CT pairs are being randomly selected at each epoch within the training dataset. To achieve independence between the style and content of an image, the perceptual loss was used [38]. The perceptual loss was used to generate the sCT by preserving the content of the CBCT and applying the style of the CT. The loss of style and content represents the projection of the input and output images into a feature space designed specifically to capture information. The content features of the pre-trained neural network capture important semantic information of the images, including their shapes, textures, and the objects in the images. By minimizing the difference between these content features for input and output images, the aim is to produce results that preserve the overall visual content of the image while modifying other aspects, such as style or color. Content loss is a measure of the similarity of high-level, multi-scale content between input images and target images.

The content loss is defined as follows:

$$\mathcal{L}_{cont}^{\phi_j}(sCT, CBCT) = \|\phi_j(sCT) - \phi_j(CBCT)\|_2^2 \quad (1)$$

where  $\phi_j(CBCT)$  is the activation at the  $j$ th layer for the CBCT

The definition of style loss is outlined as follows:

$$\text{Gram}_j^\phi(I) = \frac{\psi\psi^T}{C_j^*H_j^*W_j} \quad (2)$$

$$\psi \text{ being the flattened matrix } \phi_j(I) \text{ of size } : C_j^*H_j^*W_j \quad (3)$$

$$\mathcal{L}_{style}^{\phi_j}(sCT, CT) = \|\text{Gram}_j(sCT) - \text{Gram}_j(CT)\|_2^F \quad (4)$$

The use of style in feature maps differs from their content; instead of directly employing the feature maps, the square Frobenius norm ( $\|\dots\|_2^F$ ) between the Gram matrix  $\text{Gram}_j^\phi$  of sCT features maps and the target CT features maps were computed. The Gram matrix (Eq. 2) captures information about characteristics that tend to activate together. As demonstrated by Li [36], minimizing the Gram matrices of two images is equivalent to minimizing a specific form of maximum mean discrepancy. Consequently, this process can be interpreted as aligning the distribution of the content image with that of the style image. The perceptual loss  $\mathcal{L}_{perceptual}^{\phi_j}$  is the weighted sum of these two functions ( $\mathcal{L}_{cont}^{\phi_j}$  and  $\mathcal{L}_{style}^{\phi_j}$ ):

$$\mathcal{L}_{perceptual}^{\phi_j}(sCT, CBCT, CT) = \alpha^* \mathcal{L}_{cont}^{\phi_j}(sCT, CBCT) + \beta^* \mathcal{L}_{style}^{\phi_j}(sCT, CT) \quad (5)$$

where  $\alpha$  and  $\beta$  are the weighting factors

Due to the memory limitations of conventional perceptual loss, which was based on the VGG network, a novel perceptual loss was formulated using the ConvNext-tiny architecture referred to as Content and Style Representation for Enhanced Perceptual synthesis (CREPs) loss [39]. The ConvNext network, as detailed in the work by Liu et al. [40], mainly uses the ConvNeXt module, inspired by both convolutional neural networks and Vision Transformer (ViT) architectures [41]. ConvNext surpasses the capabilities of advanced Transformer-based models while adhering to a purely convolutional design. It adopts the standard ResNet-50 architecture [42] and refines it to align more closely with the design principles of ViT. The ConvNext-Tiny network has fewer parameters and better accuracy on object recognition tasks (28.6 M parameters and 82.52 Acc@1) than VGG-19 (143.7 M params and 72.376 Acc@1). It was used to compute a new PL with lower computational cost.

#### 2.3.2. Unsupervised MRI-to-CT synthesis

For the MRI-to-CT synthesis, a 3D method similar to that used for the CBCT approach was employed. The network was trained on CT and CBCT of two distinct patients. MRI/CT pairs are being randomly selected at each epoch within the training dataset. To impose constraints on the generator's output, only the style term of the CREPs loss was used, comparing the  $sCT_{reg}$  with a randomly selected CT image from the database.

This decision was made due to noticeable distinctions in content between MRI and CT modalities, which resulted in the inability to generate the desired output when minimizing content loss from MRI. It underscores the importance of capturing and preserving the distinctive stylistic features during the process of generating  $sCT_{reg}$  images. The unsupervised synthesis model was trained for 400 epochs with a learning rate set at 0.0002. These settings were empirically chosen. The trained model selected for use was the one obtained at the last epoch. To ensure the convergence of the synthesis network on each  $sCT_{reg}$ , a visual evaluation of the  $sCT_{reg}$  was conducted before registration.

### 2.4. Deformable image registration methods

To register CBCT/CT and MRI/CT pairs, the trained models discussed in previous sections are used for inference, resulting in the generation of an sCT and subsequent monomodal registration. This approach was compared with two multimodal techniques.

#### 2.4.1. DIR with free-form deformation (FFD)

To evaluate and compare the performance of the novel sCT-based registration approach with conventional iconic non-rigid methods, a point of reference was established using an FFD-based DIR method implemented with the Elastix library [43]. This method is one of the most used in the literature [12]. The FFD was computed using the adaptive stochastic gradient descent optimizer and normalized mutual information (NMI) similarity measure following a 4-step multi-resolution scheme, where the image resolution was divided by factors of 8, 4, 2, and 1. The final grid spacing was defined as  $14.0 \times 14.0 \times 14.0$  mm.

#### 2.4.2. DIR with VoxelMorph (VXM)

The accuracy of the proposed non-rigid registration approach was assessed by employing the VoxelMorph network [44] for performance comparison. The VoxelMorph network was used to estimate the deformation vector field (DVF) between the planning CT and daily CBCT images for each patient.

To assess the feasibility of eliminating the generation step, this approach minimizes the loss of content ( $\mathcal{L}_{CREPcontent}$ ) by registering CBCT and CT. The content feature space used in this method serves to represent the patient's anatomy in a simplified manner. Consequently, minimizing the representation of our two images within this space ensures the preservation of crucial anatomical information relating to the patient. The application of this method has been limited to CBCT, as the challenge of minimizing content loss in MRI/CT pairs hinders optimal registration, as mentioned in section 2.3.2. The disparity in content, involving distinct aspects of anatomy and physiology, between these two modalities poses a significant obstacle to effective implementation.

VoxelMorph, a weakly supervised U-Net architecture [45], was employed to generate a DVF by leveraging the inputs from both the encoder and decoder. Additionally, the estimated DVF was regularized by applying a diffusion regularizer to the spatial gradients, resulting in a smooth diffeomorphic transformation.

The loss function  $L_{VXM}$  considered in the VoxelMorph model was:

$$\mathcal{L}_{VXM} = \alpha^* \mathcal{L}_{CREPcontent}(CT, sCT) + \lambda^* \mathcal{L}_{smooth}(\phi) \quad (6)$$

where CT and sCT were the fixed and moving images and  $\phi$  was the DVF.

$\mathcal{L}_{smooth}$  was the regularization function to encourage a smooth DVF.  $\mathcal{L}_{CREPcontent}$  [39] was derived from the perceptual loss introduced by Johnson et al. [38].

The weighting parameter values, selected empirically, were:  $\alpha = 1$  and  $\lambda = 4$ . To train the VoxelMorph networks, the data were randomly separated into three folds of 50% (11 patients), 15% (3 patients), and 35% (9 patients) images, which were successively used as training, validation, and test datasets. The intensities of the images (CT and CBCT) were cropped around  $-1024$  and  $1575$ , and linearly normalized to the range  $[0, 1]$ . The VoxelMorph models were trained using the Adam optimizer with a learning rate scheduler starting at a value of  $0.0005$  and multiplied by  $0.7$  every  $10$  epochs.

#### 2.4.3. DIR through sCT generation

The novelty proposed in our study was to translate multi-modal image registration into mono-modal registration by generating an  $sCT_{reg}$  from MRI or CBCT. As in Liu et al. [9], the  $sCT_{reg}$  images were generated from the MRI or CBCT, and then the corresponding  $sCT_{reg}$  and CT images were registered. sCTs were generated in an unsupervised manner as described in the previous section.

For the registration task, the FFD method based on B-Splines, as implemented in the Elastix library [43], was employed, using identical hyperparameters to those detailed in Section 2.4.1. The only difference was the adoption of a monomodal similarity metric (normalized cross-correlation instead of the NMI), which arises from the simplification of the registration problem into a mono-modal scenario.

For CBCT/CT registration, the intensities of CT and  $sCT_{reg}$  were not

modified. However, to aid the convergence of the MRI/CT registration, a density assignment was performed in the OARs (bones, bladder, prostate, rectum) guiding the registration. Furthermore, to minimize potential edge effects induced by large deformations, the estimation of the DVF was performed in a distinct space (12 cm above and below the center of the prostate) compared to the space defined for  $sCT_{eval}$  generation (8 cm above and below the center of the prostate).

### 2.5. Registration evaluation

#### 2.5.1. Dice similarity coefficient (DSC)

To quantify the accuracy of the DIR methods, the Dice Similarity Coefficient (DSC) was calculated. This metric was computed by comparing the deformed contours to the reference contours for the rectum, bladder, and prostate. DSC was computed as follows:

$$DSC = 2 \frac{|A \cap B|}{|A| + |B|} \quad (7)$$

In this context, A and B represent the two segmentation datasets under consideration, which are the planning and deformed contours. To assess the extent of deformations within the database, the DSC after rigid registration was evaluated. To evaluate the accuracy of the DIR methods, a qualitative assessment was carried out. The evaluation was performed for each registration method: rigid registration, BSplines with and without an intermediary  $sCT_{reg}$  synthesis, and the VoxelMorph method (only for dataset 2).

This assessment involved the presentation of deformed images and masks along with their corresponding DVF for three distinct cases: the worst, median, and best scenarios, determined based on the average DSC across three anatomical structures.

#### 2.5.2. MRI-to-CT synthesis to evaluate the registration impact on supervised sCT generation accuracy

To evaluate the MRI/CT registration practically, a voxel-wise comparison of an L1 norm known as the Mean Absolute Error (MAE) between different  $sCT_{eval}$ /sCTs was performed.

The MAE in HU is calculated as:

$$MAE = \frac{1}{n} \sum_i i = 1^n |sCT_i - CT_i| \quad (8)$$

where  $i$  represents the summation index corresponding to each voxel of the image.

Images were preprocessed as described in Section 2.2, then divided into training, validation, and test sets. For training, 40 image pairs were selected (20 from each center), 4 (2 + 2) were selected for validation, and 16 (8 + 8) for the test. As reported in [3,46], the results obtained by a network trained in a monocenter or multicenter manner are similar, highlighting the network's ability to generalize.

The MRI-to-CT generation was performed with a supervised 3D conditional GAN using a ResNet-9blocks as the generator and a 707,070 PatchGAN as the discriminator. The loss function for generators was the VGG-16 perceptual loss [38], and the Binary Cross Entropy was used for discriminators. Models were trained for 400 epochs, and the best model according to Mean Square Error on the validation set was selected. Finally, a 3-fold cross-validation was performed to evaluate the stability of the synthesis.

The comparison was performed according to six cases: Cases 1–3 with test data rigidly registered for training and for the test, using respectively rigidly registered data, data registered using BSpline with MRI and CT, and data registered using BSpline with  $sCT_{reg}$  and CT. Cases 4–6 used the same training data, and test data registered by the best registration according to the DSC.

#### 2.5.3. Computation

The DL based methods were implemented in Python 3.8 using

PyTorch 1.12 with CUDA 11.7. The models were trained and tested on an NVIDIA RTX A6000 with 48 GB of VRAM. No over-fitting was observed (i.e. no increase of the loss on the validation data).

### 3. Results

#### 3.1. CBCT/CT registration results

Table 2 summarizes the DSC metrics for body, prostate, bladder, and rectum for the rigid method and 3 DIR methods.

The DSC obtained after the rigid registration is the smallest for all structures (TV and OARs). By using  $\mathcal{S}CREP_{content}$  to guide the deformation estimation, the VoxelMorph method provided better accuracy for the prostate and rectum. Compared to all other registration methods, VXM provided significantly better DSC for the bladder. Furthermore, when using  $sCT_{reg}$  instead of CBCT for multimodal registration, it outperformed rigid registration and other deformable approaches.

Fig. 2 shows the planning CT, original CBCT, and the deformed CBCTs for each method for three patients corresponding to the worst, median, and best DSC results.

#### 3.2. MRI/CT registration results

Table 3 summarizes the DSC metrics for the contours of the body, bladder, prostate, and rectum according to the registration method: rigid, the BSpline-based method with MRI and CT images, and the BSpline-based method with  $sCT_{reg}$  and CT images.

The highest DSC was obtained with the BSpline-based registration method using an intermediary  $sCT_{reg}$  for all three organs (bladder, prostate, and rectum). It reached 0.95 for bladders, 0.90 for prostates, and 0.87 for rectums for the first center, and 0.91, 0.87, and 0.84, respectively, for the second center.

To further evaluate the impact of the registration on the supervised DL MRI-to-CT synthesis, a study was performed using a 3D supervised conditional GAN (cGAN). The multicenter study was conducted on the data used in both the training and test cohorts.

Table 4 presents the results of the MRI-to-CT synthesis under six distinct scenarios. In Cases 1–3, the test data were rigidly registered for training, using rigidly registered data, data registered using BSpline with MRI and CT, and data registered using BSpline with  $sCT_{reg}$  and CT, respectively. In contrast, Cases 4–6 used identical training data, but the test data were registered using the optimal method based on DSC: BSpline using an intermediary  $sCT_{reg}$  instead of MRI.

The results reveal that the most precise Mean Absolute Error (MAE) was achieved when training with BSpline-registered data, employing an intermediary  $sCT_{reg}$  ( $39.2 \pm 5.9$  HU for C1 and  $59.6 \pm 9.3$  HU for C2). In contrast, less accurate synthesis occurred when both training and test data were rigidly registered ( $55.3 \pm 11.2$  HU for C1 and  $80.6 \pm 20.2$  HU for C2). Accurate evaluation of  $sCT_{eval}$  requires precise non-rigid registration.

Furthermore, across the cGAN models trained with three different cohorts registered with different methods, superior outcomes were consistently obtained when the test was conducted with data registered using BSpline and an intermediary  $sCT_{reg}$ . Notably, superior results were consistently observed for center C1.

**Table 2**

Registration results (mean DSC  $\pm$  std) on the 341 CBCTs according to the structure and the methodology.

Method	DSC			
	Body	Bladder	Prostate	Rectum
Rigid	0.90 $\pm$ 0.03	0.58 $\pm$ 0.15	0.75 $\pm$ 0.08	0.65 $\pm$ 0.09
VoxelMorph	0.96 $\pm$ 0.01	<b>0.90 <math>\pm</math> 0.05</b>	0.78 $\pm$ 0.06	0.72 $\pm$ 0.11
BSpline CBCT/CT	<b>0.98 <math>\pm</math> 0.01</b>	0.66 $\pm$ 0.16	0.78 $\pm$ 0.05	0.66 $\pm$ 0.08
BSpline sCT/CT	<b>0.98 <math>\pm</math> 0.01</b>	0.85 $\pm$ 0.04	<b>0.85 <math>\pm</math> 0.05</b>	<b>0.75 <math>\pm</math> 0.04</b>

### 4. Discussion

This study introduced an innovative approach to multimodal deformable image registration, enhancing accuracy. The proposed registration method, using an unsupervised generated intermediary  $sCT_{reg}$ , provides better registration results of CT with both CBCT and MRI in the pelvic area. These results were emphasized based on the Dice Similarity Coefficient (DSC), aligning with the conclusions drawn from the supervised  $sCT_{eval}$  generation procedure.

Another objective of the registration process was to perform a rigorous voxel-wise assessment of  $sCT_{eval}$ . Our method outperformed classical multimodal registration using Elastix based on B-splines (refer to Table 2 and Table 3). Furthermore, the evaluation of the second  $sCT_{eval}$  multicenter generation demonstrates that the rigidly registered training dataset leads to the worst  $sCT_{eval}$  results. This underscores the importance of registration quality not only for a less biased evaluation of  $sCT_{eval}$  but also for ensuring higher-quality training data. Consequently, the quality of registration has consequences for the convergence of the model and the accuracy of a supervised synthesis of  $sCT_{eval}$ , as mentioned by Florkow et al. [47]. Finally, all models were accurate for both centers of the training cohort, highlighting the network's ability to generalize.

One limitation associated with this registration approach is its reliance on the accuracy of the  $sCT_{reg}$ . Using an unsupervised generation method introduces potential variability in the quality of the  $sCT_{reg}$ , consequently influencing the overall registration accuracy. Even though preprocessing aims to reduce noise in MRI and artifacts in CBCTs, the registration process may still be affected by noise or artifacts.

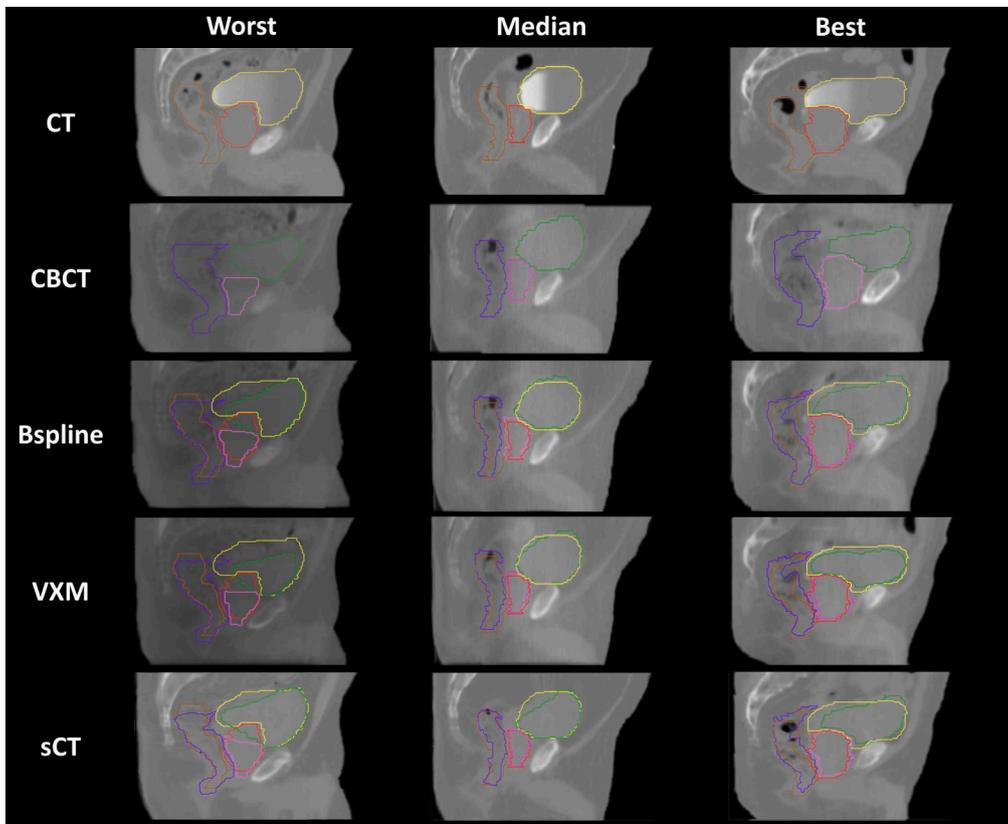
For the CBCT/CT database, transitioning from multimodal to monomodal registration provided the best scores for three out of four structures. The Dice Similarity Coefficient (DSC) achieved on the bladder with the VXM method based on CREPs loss outperformed our method. These superior performances can be attributed to the challenges posed by heterogeneity and contrast variations in CBCTs, leading to suboptimal reconstructions and artifacts in the sCT for bladders. Consequently, iconic registration faces difficulties, especially when compared to its performance for other organs. In contrast, the content loss employed in the VXM method enables effective preservation of the bladder's shape. Even though DSC results are worse than those obtained with the intermediary sCT method, it represents a first step toward accurate multimodal registration for CBCT.

It is crucial to note that existing DL registration methods place significant limitations on image input sizes and the type of deformation employed, predominantly based on optical flow. In particular, DL methods for FFD applied to medical problems are rare, highlighting the need for more versatile and adaptive approaches in this field. The DSC of the proposed method reached 0.75 on the rectum. The other DSCs are clinically acceptable ( $\geq 0.8$ ) [48,49]. These results are among the best in the literature when the registration is not guided with contours [13].

However, it is crucial to highlight that the sCT-based MRI/CT registration approach adopted in this study is hybrid, incorporating both iconic and contour-guided elements. Despite the challenges, the results obtained for the target (DSC = 0.85 for CBCT, DSC = 0.90 for MRI) and OARs are considered among the best in the existing literature [19,50,51]. It's noteworthy that these references primarily rely on iconic information for registration, limiting the scope of direct comparison due to the hybrid nature of the methodology employed in this study.

Despite the increasing interest in unsupervised generation algorithms to bypass the requirement for registration, supervised generation methods have so far demonstrated superior accuracy [4]. Additionally, although voxel-wise assessment methods are emerging [52], their adoption by the community remains limited [13].

Moreover, in radiotherapy where CT is still the standard imaging modality for planning, the daily acquisition of CBCT or MRI (with the



**Fig. 2.** CBCT/CT registration results: Row 1: CT images, Row 2: CBCT images, Rows 3–4: registered CBCT according to the registration method, respectively BSpline and VoxelMorph, and Row 5: the registered  $sCT_{reg}$ . Columns represent three different cases of registration: the worst, the median case, and the best case. Delineations are added to the images: bladder (CT in yellow, registered CBCT in green), prostate (CT in red, registered CBCT in pink), and rectum (CT in brown, registered CBCT in purple). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

**Table 3**

Registration results (mean DSC  $\pm$  std) on the 60 CT/MRI pairs are presented according to the structure (columns): body, bladder, prostate, rectum, the center (C1 (1.5 T MRI) or C2 (0.35 T MR-Linac)), and the registration methodology (rows): rigid, BSpline with MRI/CT, BSpline with  $sCT_{reg}/CT$ .

Center	Registration Method	DSC			
		Body	Bladder	Prostate	Rectum
C1	Rigid	0.98 $\pm$ 0.01	0.81 $\pm$ 0.12	0.78 $\pm$ 0.12	0.76 $\pm$ 0.08
	BSpline MRI/CT	<b>0.99 <math>\pm</math> 0.01</b>	0.93 $\pm$ 0.06	0.87 $\pm$ 0.04	0.84 $\pm$ 0.04
	BSpline $sCT/CT$	0.98 $\pm$ 0.01	<b>0.95 <math>\pm</math> 0.02</b>	<b>0.90 <math>\pm</math> 0.03</b>	<b>0.87 <math>\pm</math> 0.03</b>
C2	Rigid	0.95 $\pm$ 0.02	0.65 $\pm$ 0.16	0.75 $\pm$ 0.11	0.69 $\pm$ 0.10
	BSpline MRI/CT	<b>0.97 <math>\pm</math> 0.01</b>	0.90 $\pm$ 0.04	0.84 $\pm$ 0.06	0.78 $\pm$ 0.08
	BSpline $sCT/CT$	0.95 $\pm$ 0.01	<b>0.91 <math>\pm</math> 0.04</b>	<b>0.87 <math>\pm</math> 0.04</b>	<b>0.84 <math>\pm</math> 0.04</b>

growing significance of MR linacs) highlights the imperative for multimodal registration. The approach involves transforming the complex multimodal registration process into a more manageable monomodal registration by generating sCTs. The unsupervised synthesis algorithm based on the CREPs loss used in this process yields precise sCTs, facilitating the registration with CT.

Individuals with substantial differences in bladder or rectum volumes between imaging modalities are still a problem in the registration process. A specific case highlighted a patient from the MRI database with a bladder volume of 1200 cm<sup>3</sup> on the CT scan and 250 cm<sup>3</sup> on the MRI. Such significant disparities in volumes required extensive registration adjustments, resulting in major deformations in the images and consequently depicting unrealistic anatomies. Furthermore, an additional challenge arose from the presence of gas in the rectum within one imaging modality. This circumstance introduced a notable discrepancy between the two volumes, potentially giving rise to artifacts in the  $sCT$  or introducing bias in the evaluation process. These challenges were exacerbated by substantial differences in HU for the CT and CBCT or

intensities in the MRI, further emphasizing the complexity of the multimodal registration process in the presence of diverse anatomical variations and imaging modalities.

While this method has been assessed on two multicenter datasets, it is advisable to further evaluate its accuracy on other datasets to validate it and ensure a comprehensive and robust evaluation. Finally, current challenges in registration for radiation therapy are described in Hussein et al.'s study [53]. The study emphasizes the critical and routine requirement for clinicians in utilizing DL algorithms for registration purposes. It underlines the potential of DL algorithms in addressing these needs. However, it also emphasizes the necessity for a dependable quality analysis to instill trust in the reliability of these new tools.

This study highlights the need for more accurate multimodal registration, specifically in the pelvic region involving CBCT/CT and MRI/CT. The versatility of this approach extends beyond the pelvic area, showcasing potential applications in various anatomical regions. However, it's crucial to acknowledge the primary constraint, which revolves around the convergence challenges associated with the unsupervised

**Table 4**

Mean Absolute Error (MAE)  $\pm$  std. (in HU) between  $sCT_{eval}$  and reference CT according to the registration in the train and in the test. The six cases are as follows: Cases 1–3 involve test data rigidly registered for both training and testing, using respectively rigidly registered data, data registered using BSpline with MRI and CT, and data registered using BSpline with  $sCT_{reg}$  and CT. Cases 4–6 use the same training data, with test data registered using the best registration according to the DSC.

Registration	test data, rigid registration			test data, registration BSpline sCT/CT		
	Rigid	BSpline MRI/CT	BSpline sCT/CT	Rigid	BSpline MRI/CT	BSpline sCT/CT
MAE for C1	55.3	46.5 $\pm$ 6.9	45.4 $\pm$ 7.7	42.1 $\pm$ 5.7	41.2 $\pm$ 5.1	<b>39.2 <math>\pm</math> 5.9</b>
	$\pm$ 11.2					
MAE for C2	80.6	69.1 $\pm$ 14.7	67.5 $\pm$ 11.2	69.6 $\pm$ 16.3	63.1 $\pm$ 12.0	<b>59.6 <math>\pm</math> 9.3</b>
	$\pm$ 20.2					

$sCT_{reg}$  generation model. Furthermore, the proposed method presents broader applicability, encompassing areas such as motion detection, daily registration for tumor monitoring, and more.

## 5. Conclusion

In conclusion, the incorporation of intermediary synthetic images, generated through the implementation of an unsupervised deep learning model, stands out as a key advancement that significantly improves the precision of multimodal registration. The VoxelMorph registration results suggest that integrating novel non-linear multimodal deep learning metrics may enhance registration performance compared to conventional metrics. This study not only establishes the benefits of this approach but also underlines the critical importance of registration accuracy, particularly in the context of supervised sCT generation. The findings of this research pave the way for broader application of the proposed approach, extending its utility beyond the anatomical regions and imaging modalities examined in the current study.

## CRediT authorship contribution statement

**Cédric Hémon:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization. **Blanche Texier:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization. **Hilda Chourak:** Writing – review & editing, Data curation. **Antoine Simon:** Writing – review & editing, Data curation. **Igor Bessières:** Writing – review & editing, Data curation. **Renaud de Crevoisier:** Writing – review & editing, Supervision, Project administration, Funding acquisition. **Joël Castelli:** Writing – review & editing, Supervision. **Caroline Lafond:** Writing – review & editing, Validation, Supervision. **Anaïs Barateau:** Writing – review & editing, Validation, Supervision, Investigation. **Jean-Claude Nunes:** Writing – review & editing, Validation, Supervision, Project administration, Funding acquisition.

## Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Cedric Hemon reports financial support was provided by Elekta AB. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

## Data availability

The authors do not have permission to share data.

## Acknowledgement

This research was supported by a PhD scholarship Grant from Elekta AB and a PhD scholarship Grant from University of Rennes. The present work was also funded by CominLabs with the CEMMTAUR project 2022.

## References

- [1] B. Schaly, J.A. Kempe, G.S. Bauman, J.J. Battista, J.V. Dyk, Tracking the dose distribution in radiation therapy by accounting for variable anatomy, *Phys. Med. Biol.* 49 (5) (2004) 791, <https://doi.org/10.1088/0031-9155/49/5/010>.
- [2] C. Hemon, B. Rigaud, A. Barateau, F. Tilquin, V. Noblet, D. Sarrut, P. Meyer, J. Bert, R. De Crevoisier, A. Simon, Contour-guided deep learning based deformable image registration for dose monitoring during cbct-guided radiotherapy of prostate cancer, *J. Appl. Clin. Med. Phys.* (2023) e13991, <https://doi.org/10.1002/acm2.13991>.
- [3] B. Texier, C. Hémon, P. Lekieffre, E. Collot, S. Tahri, H. Chourak, J. Dowling, P. Greer, I. Bessieres, O. Acosta, et al., Computed tomography synthesis from magnetic resonance imaging using cycle generative adversarial networks with multicenter learning, *Phys. Imag. Radiat. Oncol.* 28 (2023) 100511, <https://doi.org/10.1016/j.phro.2023.100511>.
- [4] M.F. Spadea, M. Maspero, P. Zaffino, J. Seco, Deep learning based synthetic-ct generation in radiotherapy and pet: a review, *Med. Phys.* 48 (11) (2021) 6537–6566, <https://doi.org/10.1002/mp.15150>.
- [5] M. Boulanger, J.-C. Nunes, H. Chourak, A. Largent, S. Tahri, O. Acosta, R. de Crevoisier, C. Lafond, A. Barateau, Deep learning methods to generate synthetic CT from MRI in radiotherapy: a literature review, *Phys. Med.* 89 (2021) 265–281, <https://doi.org/10.1016/j.ejmp.2021.07.027>.
- [6] A. Kazerouni, E.K. Aghdam, M. Heidari, R. Azad, M. Fayyaz, I. Hacihaliloglu, D. Merhof, Diffusion models for medical image analysis: A comprehensive survey, *arXiv*, 2022, <https://doi.org/10.48550/arXiv.2211.07804>.
- [7] J. Chen, S. Chen, L. Wee, A. Dekker, I. Bermejo, Deep learning based unpaired image-to-image translation applications for medical physics: a systematic review, *Phys. Med. Biol.* (2023), <https://doi.org/10.1088/1361-6560/acba74>.
- [8] Z. Tang, P.-T. Yap, D. Shen, A new multi-atlas registration framework for multimodal pathological images using conventional monomodal normal atlases, *IEEE Trans. Image Process.* 28 (5) (2019) 2293–2304, conference Name: IEEE Transactions on Image Processing, <https://doi.org/10.1109/TIP.2018.2884563>.
- [9] X. Liu, D. Jiang, M. Wang, Z. Song, Image synthesis-based multi-modal image registration framework by using deep fully convolutional networks, *Med. Biol. Eng. Comput.* 57 (5) (2019) 1037–1048, <https://doi.org/10.1007/s11517-018-1924-y>.
- [10] B. Rigaud, A. Simon, J. Castelli, C. Lafond, O. Acosta, P. Haigrón, G. Cazoulat, R. de Crevoisier, Deformable image registration for radiation therapy: principle, methods, applications and evaluation, *Acta Oncol.* 58 (9) (2019) 1225–1237, <https://doi.org/10.1080/0284186X.2019.1620331>.
- [11] S. Oh, S. Kim, Deformable image registration in radiation therapy, *Radiat. Oncol. J.* 35 (2) (2017) 101, <https://doi.org/10.3857/roj.2017.00325>.
- [12] C. Dossun, C. Niederst, G. Noel, P. Meyer, Evaluation of dir algorithm performance in real patients for radiotherapy treatments: a systematic review of operator-dependent strategies, *Phys. Med.* 101 (2022) 137–157, <https://doi.org/10.1016/j.ejmp.2022.08.011>.
- [13] L. Nenoff, F. Amstutz, M. Murr, B. Archibald-Heeren, M. Fusella, M. Hussein, W. Lechner, Y. Zhang, G. Sharp, E.V. Osorio, Review and recommendations on deformable image registration uncertainties for radiotherapy applications, *Phys. Med. Biol.* 68 (24) (2023) 24TR01, <https://doi.org/10.1088/1361-6560/ad0d8a>.
- [14] D. Sengupta, P. Gupta, A. Biswas, A survey on mutual information based medical image registration algorithms, *Neurocomputing* 486 (2022) 174–188, <https://doi.org/10.1016/j.neucom.2021.11.023>.
- [15] F. Maes, D. Vandermeulen, P. Suetens, Medical image registration using mutual information, *Proc. IEEE* 91 (10) (2003) 1699–1722, <https://doi.org/10.1109/JPROC.2003.817864>.
- [16] Z. Gao, B. Gu, J. Lin, Monomodal image registration using mutual information based methods, *Image Vis. Comput.* 26 (2) (2008) 164–173, <https://doi.org/10.1016/j.imavis.2006.08.002>.
- [17] J. Zou, B. Gao, Y. Song, J. Qin, A review of deep learning-based deformable medical image registration, *Front. Oncol.* 12 (2022).
- [18] G. Wu, M. Kim, Q. Wang, Y. Gao, S. Liao, D. Shen, Unsupervised deep feature learning for deformable registration of MR brain images, in: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013, Lecture Notes in Computer Science*, Springer, Berlin, Heidelberg, 2013, pp. 649–656, [https://doi.org/10.1007/978-3-642-40763-5\\_80](https://doi.org/10.1007/978-3-642-40763-5_80).
- [19] Y. Liu, W. Wang, Y. Li, H. Lai, S. Huang, X. Yang, Geometry-consistent adversarial registration model for unsupervised multi-modal medical image registration, *IEEE J. Biomed. Health Inform.* 27 (7) (2023) 3455–3466, <https://doi.org/10.1109/JBHI.2023.3270199>.
- [20] R. Han, C.K. Jones, J. Lee, X. Zhang, P. Wu, P. Vagdari, A. Uneri, P.A. Helm, M. Luciano, W.S. Anderson, J.H. Siewerdsen, Joint synthesis and registration

- network for deformable MR-CBCT image registration for neurosurgical guidance, *Phys. Med. Biol.* 67 (2022) 125008, <https://doi.org/10.1088/1361-6560/ac72ef>.
- [21] L. Deng, H. Sun, J. Wang, S. Huang, X. Yang, A novel unsupervised mri synthetic ct image generation framework with registration network, *Comp. Mater. Continua* 77 (2) (2023), <https://doi.org/10.32604/cmc.2023.039062>.
- [22] L. Zhou, X. Ni, Y. Kong, H. Zeng, M. Xu, J. Zhou, Q. Wang, C. Liu, Mitigating misalignment in mri-to-ct synthesis for improved synthetic ct generation: an iterative refinement and knowledge distillation approach, *Phys. Med. Biol.* 68 (24) (2023) 245020.
- [23] D. Wang, J. Liu, X. Fan, R. Liu, Unsupervised misaligned infrared and visible image fusion via cross-modality image generation and registration, *arXiv*, 2022, <https://doi.org/10.48550/arXiv.2205.11876>.
- [24] M.M. Saad, R. O'Reilly, M.H. Rehmani, A survey on training challenges in generative adversarial networks for biomedical image analysis, *Artif. Intell. Rev.* 57 (2) (2024) 19, <https://doi.org/10.1007/s10462-023-10624-y>.
- [25] D. Hong, B. Zhang, X. Li, Y. Li, C. Li, J. Yao, N. Yokoya, H. Li, P. Ghamisi, X. Jia, A. Plaza, P. Gamba, J.A. Benediktsson, J. Chanussot, SpectralGPT: spectral remote sensing foundation model, *IEEE Trans. Pattern Anal. Mach. Intell.* (2024), <https://doi.org/10.1109/TPAMI.2024.3362475>, 1–18ArXiv:2311.07113 [cs].
- [26] T. Chen, S. Kornblith, M. Norouzi, G. Hinton, A Simple Framework for Contrastive Learning of Visual Representations, 2020, <https://doi.org/10.48550/arXiv.2002.05709>.
- [27] W. Jiangtao, W. Xinhong, J. Xiao, Y. Bing, Z. Lei, Y. Yidong, MRI to CT synthesis using contrastive learning, 2021 IEEE International Conference on Medical Imaging Physics and Engineering (ICMIPE), 2021, pp. 1–5. doi: <https://doi.org/10.1109/ICMIPE53131.2021.9698888>.
- [28] H. Emami, M. Dong, C. Glide-Hurst, CL-GAN: Contrastive learning-based generative adversarial network for modality transfer with limited paired data, in: L. Karlinsky, T. Michaeli, K. Nishino (Eds.), *Computer Vision – ECCV 2022 Workshops*, Springer Nature, Switzerland, Cham, 2023, pp. 527–542, [https://doi.org/10.1007/978-3-031-25066-8\\_30](https://doi.org/10.1007/978-3-031-25066-8_30).
- [29] J. Wasserthal, H.-C. Breit, M.T. Meyer, M. Pradella, D. Hinck, A.W. Sauter, T. Heye, D.T. Boll, J. Cyriac, S. Yang, M. Bach, M. Segeroth, TotalSegmentator: robust segmentation of 104 anatomic structures in CT images, *Radiol.: Artif. Intell.* 5 (5) (2023) e230024, <https://doi.org/10.1148/ryai.230024>.
- [30] H. Yang, P. Qian, C. Fan, An indirect multimodal image registration and completion method guided by image synthesis, *Comput. Math. Methods Med.* 2020 (2020), <https://doi.org/10.1155/2020/2684851>, 2684851, publisher: Hindawi.
- [31] E.M.C. Huijben, M.L. Terpstra, A.J. Galapon, S. Pai, A. Thummerer, P. Koopmans, M. Afonso, M. van Eijnatten, O. Gurney-Champion, Z. Chen, Y. Zhang, K. Zheng, C. Li, H. Pang, C. Ye, R. Wang, T. Song, F. Fan, J. Qiu, Y. Huang, J. Ha, J.S. Park, A. Alain-Beaudoin, S. Bériault, P. Yu, H. Guo, Z. Huang, G. Li, X. Zhang, Y. Fan, H. Liu, B. Xin, A. Nicolson, L. Zhong, Z. Deng, G. Müller-Franzes, F. Khader, X. Li, Y. Zhang, C. Hémon, V. Boussoit, Z. Zhang, L. Wang, L. Bai, S. Wang, D. Mus, B. Kooiman, C.A.H. Sargeant, E.G.A. Henderson, S. Kondo, S. Kasai, R. Karimzadeh, B. Ibragimov, T. Helfer, J. Dafflon, Z. Chen, E. Wang, Z. Perko, M. Maspero, Generating Synthetic Computed Tomography for Radiotherapy: SynthRAD2023 Challenge Report, 2024 doi:<http://arxiv.org/abs/2403.08447>.
- [32] N.J. Tustison, B.B. Avants, P.A. Cook, Y. Zheng, A. Egan, P.A. Yushkevich, J.C. Gee, N4itk: improved n3 bias correction, *IEEE Trans. Med. Imaging* 29 (6) (2010) 1310–1320, doi:[10.1109/TMI.2010.2046908](https://doi.org/10.1109/TMI.2010.2046908).
- [33] P. Perona, J. Malik, Scale-space and edge detection using anisotropic diffusion, *IEEE Trans. Pattern Anal. Mach. Intell.* 12 (7) (1990) 629–639, <https://doi.org/10.1109/34.56205>.
- [34] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial networks, *Commun. ACM* 63 (11) (2020) 139–144, doi:[doi:10.1145/3422622](https://doi.org/10.1145/3422622).
- [35] J. Huang, J. Liao, S. Kwong, Unsupervised image-to-image translation via pre-trained StyleGAN2 network, *IEEE Trans. Multimed.* 24 (2022) 1435–1448, <https://doi.org/10.1109/TMM.2021.3065230>.
- [36] Y. Li, N. Wang, J. Liu, X. Hou, Demystifying Neural Style Transfer, 2017, <https://doi.org/10.48550/arXiv.1701.01036>.
- [37] H. Guan, M. Liu, Domain adaptation for medical image analysis: a survey, *IEEE Trans. Biomed. Eng.* 69 (3) (2022) 1173–1185, <https://doi.org/10.1109/TBME.2021.3117407>.
- [38] J. Johnson, A. Alahi, L. Fei-Fei, Perceptual Losses for Real-Time Style Transfer and Super-Resolution 9906, 2016, pp. 694–711, [https://doi.org/10.1007/978-3-319-46475-6\\_43](https://doi.org/10.1007/978-3-319-46475-6_43).
- [39] C. Hémon, V. Boussoit, B. Texier, J.-L. Dillenseger, J.-C. Nunes, Guiding Unsupervised CBCT-to-CT synthesis using Content and style Representation by an Enhanced Perceptual synthesis (CREPS) loss, in: *SynthRAD2023 Challenge*, MICCAI 2023, 2023. URL <https://hal.science/hal-04272509>.
- [40] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, S. Xie, A ConvNet for the 2020s, Mar. 2022, <https://doi.org/10.48550/arXiv.2201.03545>.
- [41] W. Chen, X. Du, F. Yang, L. Beyer, X. Zhai, T.-Y. Lin, H. Chen, J. Li, X. Song, Z. Wang, D. Zhou, A Simple Single-Scale Vision Transformer for Object Localization and Instance Segmentation, Oct. 2022. URL <http://arxiv.org/abs/2112.09747>.
- [42] K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition, Dec. 2015.
- [43] S. Klein, M. Staring, K. Murphy, M.A. Viergever, J.P.W. Pluim, Elastix: a toolbox for intensity-based medical image registration, *IEEE Trans. Med. Imaging* 29 (1) (2010) 196–205, <https://doi.org/10.1109/TMI.2009.2035616>.
- [44] G. Balakrishnan, A. Zhao, M.R. Sabuncu, J. Guttag, A.V. Dalca, VoxelMorph: a learning framework for deformable medical image registration, *IEEE Trans. Med. Imaging* 38 (8) (2019) 1788–1800, <https://doi.org/10.1109/TMI.2019.2897538>.
- [45] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, May 2015, <https://doi.org/10.48550/arXiv.1505.04597>.
- [46] S. Tahri, B. Texier, J.-C. Nunes, C. Hémon, H. Chourak, J. Le Guevelou, P. Greer, J. Dowling, O. Acosta, I. Bessieres, et al., A deep learning model to generate synthetic ct for prostate mr-only radiotherapy dose planning: a multicenter study, *Front. Oncol.* 13 (2023) 1279750, <https://doi.org/10.3389/fonc.2023.1279750>.
- [47] M.C. Florkow, F. Zijlstra, L.G. Kerkmeijer, M. Maspero, C.A. van den Berg, M. van Stralen, P.R. Seevinck, The impact of mri-ct registration errors on deep learning-based synthetic ct generation, in: *Medical Imaging 2019: Image Processing* vol. 10949, SPIE, 2019, pp. 831–837, <https://doi.org/10.1117/12.2512747>.
- [48] K.K. Brock, S. Mutic, T.R. McNutt, H. Li, M.L. Kessler, Use of image registration and fusion algorithms and techniques in radiotherapy: Report of the aapm radiation therapy committee task group no. 132, *Med. Phys.* 44 (7) (2017), <https://doi.org/10.1002/mp.12256> e43–e76.
- [49] S. Thörnqvist, J.B. Petersen, M. Høyer, L.N. Bentzen, L.P. Muren, Propagation of target and organ at risk contours in radiotherapy of prostate cancer using deformable image registration, *Acta Oncol.* 49 (7) (2010) 1023–1032, <https://doi.org/10.3109/0284186X.2010.503662>.
- [50] T. Ishida, N. Kadoya, S. Tanabe, H. Ohashi, H. Nemoto, S. Dobashi, K. Takeda, K. Jingu, Evaluation of performance of pelvic ct-mr deformable image registration using two software programs, *J. Radiat. Res.* 62 (6) (2021) 1076–1082, <https://doi.org/10.1093/jrr/rrab078>.
- [51] I.D. Kolenbrander, M. Maspero, A.A. Hendriksen, R. Pollitt, J.R. van der Voort van Zyp, C.A. van den Berg, J.P. Pluim, M.A. van Eijnatten, Deep-learning-based joint rigid and deformable contour propagation for magnetic resonance imaging-guided prostate radiotherapy, *Med. Phys.* (2024), <https://doi.org/10.1002/mp.17000>.
- [52] H. Chourak, A. Barateau, S. Tahri, C. Cadin, C. Lafond, J.-C. Nunes, A. Boue-Rafle, M. Perazzi, P. Greer, J. Dowling, et al., Quality assurance for mri-only radiation therapy: a voxel-wise population-based methodology for image and dose assessment of synthetic ct generation methods, *Front. Oncol.* 12 (2022) 968689, <https://doi.org/10.3389/fonc.2022.968689>.
- [53] M. Hussein, A. Akintonde, J. McClelland, R. Speight, C.H. Clark, Clinical use, challenges, and barriers to implementation of deformable image registration in radiotherapy—the need for guidance and qa tools, *Br. J. Radiol.* 94 (1122) (2021) 20210001, <https://doi.org/10.1259/bjr.20210001>.