RESEARCH

Synthetic CT generation for pelvic cases based on deep learning in multi-center datasets

Xianan Li¹, Lecheng Jia^{2,3*}, Fengyu Lin², Fan Chai¹, Tao Liu¹, Wei Zhang⁴, Ziquan Wei², Weiqi Xiong⁴, Hua Li², Min Zhang¹ and Yi Wang^{1*}

Abstract

Background and purpose To investigate the feasibility of synthesizing computed tomography (CT) images from magnetic resonance (MR) images in multi-center datasets using generative adversarial networks (GANs) for rectal cancer MR-only radiotherapy.

Materials and methods Conventional T2-weighted MR and CT images were acquired from 90 rectal cancer patients at Peking University People's Hospital and 19 patients in public datasets. This study proposed a new model combining contrastive learning loss and consistency regularization loss to enhance the generalization of model for multi-center pelvic MRI-to-CT synthesis. The CT-to-sCT image similarity was evaluated by computing the mean absolute error (MAE), peak signal-to-noise ratio (SNRpeak), structural similarity index (SSIM) and Generalization Performance (GP). The dosimetric accuracy of synthetic CT was verified against CT-based dose distributions for the photon plan. Relative dose differences in the planning target volume and organs at risk were computed.

Results Our model presented excellent generalization with a GP of 0.911 on unseen datasets and outperformed the plain CycleGAN, where MAE decreased from 47.129 to 42.344, SNRpeak improved from 25.167 to 26.979, SSIM increased from 0.978 to 0.992. The dosimetric analysis demonstrated that most of the relative differences in dose and volume histogram (DVH) indicators between synthetic CT and real CT were less than 1%.

Conclusion The proposed model can generate accurate synthetic CT in multi-center datasets from T2w-MR images. Most dosimetric differences were within clinically acceptable criteria for photon radiotherapy, demonstrating the feasibility of an MRI-only workflow for patients with rectal cancer.

Keywords Synthetic CT, Generative adversarial nets, Contrastive learning, Multi center

*Correspondence: Lecheng Jia lecheng,Jia@cri-united-imaging.com Yi Wang wang_yi@bjmu.edu.cn ¹Peking University People's Hospital, Beijing, China ²Shenzhen United Imaging Research Institute of Innovative Medical Equipment, Shenzhen, China ³Zhejiang Engineering Research Center for Innovation and Application of Intelligent Radiotherapy Technology, Wenzhou, China ⁴Shanghai United Imaging Healthcare Co., Ltd. Shanghai, Shanghai, China

© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0./. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.





Open Access

Introduction

Recently, Magnetic resonance only (MR-only) radiotherapy has become a common research focus since it first raised, due to the superior soft-tissue contrast of MR images compared to computed tomography (CT) [1, 2]. Moreover, MR-only radiotherapy avoids additional radiation [3]. However, MR images do not contain electron density information, which is necessary for dose calculation in MR-only radiotherapy [4]. The standard solution is to generate synthetic CT (sCT) from MR images. Lately, with the development of deep learning, it has also shown great potential in the generation of sCT. Deep learning methods use large-scale image samples more efficiently to learn complex MR-to-CT mapping than conventional methods [5]. Besides, sCT can be generated during the model inference phase in just a few seconds, enabling faster model deployment [5].

The 2D Deep Convolutional Neural Network (DCNN) were first applied to the generation of sCT from MR images in the brain [5]. With the proposal of Generative Adversarial Network (GAN) [6], the application of GAN to generate images has become mainstream. For example, Conditional Generative Adversarial Network (CGAN) were used to solve the generation of sCT in the abdomen and brain [7–11]. Nevertheless, DCNN and CGAN require strictly paired data for training, which severely limits their application and increases the difficulty of data collection [12]. The CycleGAN was proposed to train the model on unpaired data through cycle consistency [13]. It was trained on unpaired data and achieved comparable results of paired data in the brain [14]. However, cycleconsistency is often too restrictive which assumes that the relationship between the two domains is a bijection in CycleGAN [15]. The contrastive learning (CL) loss was firstly utilized to enlarge the mutual information between the same location of input and synthetic images in CUT followed by some improvements, such as NEGCUT and F-LSeSim [16–18]. However, the semantic relationship between image patches is ignored and all negative image patches are considered equal probability patches.

Currently, there are some papers discussing the generalization of models in generation tasks [19–21], but a few researches focus on improving the generalization performance in multi-central datasets. This is commonly achieved by domain adaptation or using strong data augmentation in segmentation tasks [22–25]. However, domain adaptation needs to retrain the model using unseen domain data, reducing the practicability of the model [24]. Strong data augmentation, such as color augmentation, could destroy the distribution of original data and cannot be applied directly in generation tasks.

In this study, a novel framework was developed to enhance the generalization of the model in multi-central datasets using the consistency regularization learned from semi-supervised learning [26, 27]. The improved contrastive learning loss in consideration of semantic relationship was employed to enhance the structural consistency between MR and sCT.

Materials and methods

Data acquisition

The study cohort consisted of 90 patients diagnosed with rectal cancer from April 2018 to March 2021 at Peking University People's Hospital (PUPH) and 19 patients diagnosed with rectal or prostate cancer in public datasets from three different Swedish radiotherapy departments [28]. Enough patient data of rectal cancer with consistent standards can be provided in this center. One of the aim of this work is to build an MR-only workflow for rectal cancer treatment. The data acquisition parameters are shown in Table 1.

The age distribution of patients was 43–83 years in PUPH cohort. CT scanning was performed with a Philips 16-row large-aperture analog positioning machine with a flat table top. Scan parameters: 140 kV, 280 mAs, layer thickness 3 mm. MRI of the pelvis was performed using a GE Discovery MR750 3.0T MR scanner with the curved table top. The scanning sequence and parameters are as follows: High-resolution non-fat-suppressed fast recovery fast spin-echo (FRFSE) T2-weighted imaging

| Table 1 The datase | t acquisition | parameters |
|--------------------|---------------|------------|
|--------------------|---------------|------------|

| | PUPH | Public datasets | | | |
|---------------------------------------|--------------------|--------------------|--------------------------------|-----------------|--|
| | | Site 1 | Site 2 | Site 3 | |
| Number of patients MRI | 90 | 8 | 7 | 4 | |
| Manufacturer | GE Discovery MR750 | GE Discovery MR750 | Siemens Aera | GE Signa PET/MR | |
| Field strength (T) | 3 | 3 | 1.5 | 3 | |
| Spatial resolution (mm ³) | 0.67*0.67*7 | 0.875*0.875*3 | 0.875*0.875*2.5 1.1*1.1*2.5 | 0.875*0.875*2.5 | |
| CT | | | | | |
| Manufacturer | Phillips | Siemens | Toshiba | Siemens | |
| Spatial resolution (mm ³) | 1*1*3 | 0.98*0.98*3 | 1*1*2 | 0.98*0.98*2.5 | |

sequence, TR 3 200 ms, TE 85 ms, slice thickness 3 mm, slice interval 0.5 mm, a field of view 32 cm \times 26 cm.

T2-weighted MR and CT data were collected for 19 patients at three different sites in public datasets. All patients were scanned using a coil setup that not affects the outline of the patient in the radiotherapy treatment position with a flat table top.

Preprocessing

External contours of CT and MR images were generated in Treatment Planning System (TPS) (UIH, Shanghai United Imaging Healthcare Co., Ltd.). All CT and MR voxels outside the external contour were assigned to intensity of -1024 and 0, respectively. The intensities of CT images were linearly mapped from [-1024; 1500] to [-1; 1]. The intensities of MR images were clipped beyond the 95th percentile, and then the intensities were also linearly mapped to [-1; 1]. Deformable registration was performed on MR and CT images by NiftyReg open-source software [29], and the registration results were revised by an experienced physician.

The three-fold cross-validation was used in this study. The specific division of data is as follows: 30 cases were randomly selected from PUPH cohort and one center was picked from public datasets as test datasets and the rest data served as train datasets in each fold.

Network architectures

As shown in Fig. 1(a), the proposed CRGAN (Consistency Regularization Generative Adversarial Network) contained two generators and discriminators. Wherein the generator G_{CT} provides MR to CT mapping, the generator G_{MR} provides CT to MR mapping. Furthermore, the discriminator D_{MR} and D_{CT} were used to distinguish between real images and synthetic images [13].

Figure 1a) and b) show the training phase of CRGAN. In order to improve the generalization of the model, consistency regularization similar to Flexmatch was employed to optimize the G_{CT} [26], as shown in Fig. 1b). The weak and strong data augmentation was performed in the same MR image to obtain the MRw and MRs. In weak data augmentation, operations such as like flipping along the vertical direction, scaling and clip to certain size, random clip and resize, and rotation with random degree between $0 \sim 360^\circ$ were applied without changing the value distribution of the image. In strong data augmentation, MR images were further operated with color augmentation, includes the methods that will change the voxel values of images, such as altering the brightness using gamma changes, applying Gaussian filtering to the image. Then the consistency regularization loss was added to ensure that the weak and strong augmentation MR images would generate similar sCT. Figure 1c) show the inference phase of the model.

A 2.5D image was token as input of CRGAN, which contains 3 adjacent layers is extract from a 3D image. The ADAM optimization was used to minimize the loss function [30]. CRGAN was initialized using the He_normal initialization method [31].



Fig. 1 Illustration of architecture of CRGAN. (a) and (b) are the training phase of CRGAN, (c) is the inference phase of CRGAN

Generator

The Transformer module was employed in the generator of CRGAN, as shown in Fig. 2. The Transformer module can pay attention to the global connection of features compared with the convolution module [32]. There are tremendous work including imaging segmentation and translation adopting transformer structures and obtain promising performance. It is generally believed that the Transformer module is more effective than the convolution module in extracting deep features [33], so we put the Transformer module on the last layer of the encoder of the generator.

Discriminator

All the discriminator networks in CRGAN shared the same architecture, obtained by spectral normalization of the discriminator in the plain CycleGAN [13]. Spectral normalization introduces regularity constraints from the

perspective of the spectral norm of the parameter matrix of each layer of the neural network [34], so that the neural network has better insensitivity to input disturbances, thus making the training process more stable and easier to converge.

Loss function

In this study, a mixed loss function including adversarial loss, cycle consistency loss, consistency regularization loss and contrastive learning loss was used as the objective function, which is defined as follows:

$$Loss = L_{adv} + L_{cycle} + L_{consistency regulation} + L_{contrastive learning}$$

The adversarial loss (shown as L_{GAN} in Fig. 1) function optimized the generator and discriminator. For the



Fig. 2 Illustration of architecture of generator of CRGAN. IN: Instance Norm, LRelu: LeakyRelu, LN: Layer norm, FFN: Feed Forward Network

generator G_{CT} and its discriminator D_{CT} , the adversarial loss function is defined as

$$L_{adv}(G_{CT}, D_{CT}) = D_{CT}(G_{CT}(I_{MR})) + (1 - D_{CT}(I_{CT}))$$

Where $I_{\rm CT}$ and $I_{\rm MR}$ represent unpaired input CT and MR images. In the training phase, $G_{\rm CT}$ generates a synthetic CT image $G_{\rm CT}(I_{\rm MR})$ that is close to the real CT image, while $D_{\rm CT}$ is to distinguish the synthetic CT image $G_{\rm CT}(I_{\rm MR})$ from a real image $I_{\rm CT}$. Likewise, the adversarial loss functions for $G_{\rm MR}$ and $D_{\rm MR}$ are defined as

$$L_{adv}(G_{MR}, D_{MR}) = D_{MR}(G_{MR}(I_{CT})) + (1 - D_{MR}(I_{MR}))$$

The cycle-consistent loss function optimized the G_{CT} and G_{MR} , forcing the reconstructed images G_{CT} ($G_{MR}(I_{CT})$) and G_{MR} ($G_{CT}(I_{MR})$) to be the same as their input I_{CT} and I_{MR} . This loss function is defined as

$$L_{cycle} (G_{CT}, G_{MR}) = \|G_{CT} (G_{MR} (I_{CT})) - I_{CT}\| + \|G_{MR} (G_{CT} (I_{MR})) - I_{MR}\|$$

The consistency regularization loss (shown as L_{Cycle} in Fig. 1. (a)) function optimized the G_{CT} , ensuring MR images enhanced by weak and strong augmentation would generate similar sCT. This loss function is defined as

$$L_{consistency regularization} (G_{CT}) = \|G_{CT} (I_{MRw}) - G_{CT} (I_{MRs})\|$$

Contrastive learning loss

The CL loss (shown as L_{Cons} in Fig. 1. (b)) optimized the generator G_{CT} and G_{MR} . The semantic relation consistency (SRC) regularization with the decoupled contrastive learning was used [15]. SRC utilizes the semantics feature by focusing on the semantic relation between the image patches from a single image. In addition, the hard negative mining strategy is explored by exploiting the semantic relation [15]. This loss function is defined as

$$L_{CL} = \gamma_{SRC} \sum_{k=1}^{K} JSD\left(\frac{\exp\left(z_{k}^{T} z_{i}\right)}{\sum_{j=1}^{K} \exp\left(z_{k}^{T} z_{j}\right)} \middle| \left| \frac{\exp\left(w_{k}^{T} w_{i}\right)}{\sum_{j=1}^{K} \exp\left(w_{k}^{T} w_{j}\right)} \right) + \gamma_{hDCE} \mathbb{E}_{(z,w) \sim p_{ZW}} [-\log \frac{\exp(w^{T} z/\tau)}{N \mathbb{E}_{z^{-} \sim q_{T}} [\exp(w^{T} z^{-}/\tau)]}]$$

where γ_{SRC} and γ_{hDCE} are weighting parameters; JSD represents Jensen-Shannon Divergence; z_k and z_i are the corresponding embedding vectors of k-th location and i-th location patches of the input image; w_k and w_i are the corresponding embedding vectors of k-th location and i-th location patches of the synthetic image; the negative sampling is modeled by von Mises-Fisher distribution:

$$z^{-} \sim \mathbf{q}_{Z^{-}}(z^{-}; z, \gamma) = \frac{1}{N_{q}} \exp\left\{\gamma(z^{T}z^{-})\right\} p_{Z}\left(z^{-}\right)$$

where N_q is a normalization constant; γ is a hyper-parameter determining the hardness of the negative samples.

Evaluation metrics

Referring to the main stream articles of imaging translation, MAE, SNRpeak and SSIM are three commonly used metrics to measure the quality of images. MAE evaluates the voxel-vise similarity between images, SNRpeak evaluates the image quality, and the SSIM evaluates the structure similarity between two images. To evaluate the generalization performance across multi-center data, we proposed a new metric GP (Generalization Performance) based on the above metrics. To evaluating the similarity between a sCT image and real CT, their image quality, HU values, and anatomical structures similarity are most important aspects, and the metrics mentioned above can evaluate the performance from these aspects.

MAE (mean absolute error)

MAE can be used to evaluate the difference in HU values between sCT and CT images as follows:

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |CT_i - sCT_i|$$

Where the index *i* represents each voxel of the image.

SNRpeak (Peak Signal-to-noise ratio)

SNRpeak provides an objective measure of image distortion or noise level, as follows: $SNRpeak = 10^{*}log \qquad \left(\frac{MAX_{I}^{2}}{L} \right)$

$$SNTPEak = 10 \ log_{10} \left(\frac{MSE}{MSE} \right)$$

SSIM (Structural Similarity index)

SSIM analyzes the similarity between images in terms of brightness, contrast, and structure, as follows: C1 and C2 are constants.

SSIM =
$$\frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

Model's generalization analysis

We proposed a new metric GP (Generalization Performance) to assess the generalization of model on unseen datasets. It is computed as follows:

$$GP = \frac{MAE_{seen}}{MAE_{unseen}} * \frac{PSNR_{unseen}}{PSNR_{seen}} * \frac{SSIM_{unseen}}{SSIM_{seen}}$$

GP is composed of three parts, each of which reflects the model's generalization in MAE, SNRpeak and SSIM, respectively. Therefore, this indicator can comprehensively reflect the generalization performance on unseen datasets. The seen and unseen datasets are defined as the datasets trained on and test on, respectively. In such definition, seen and unseen datasets are whole datasets, not the training or testing splits from one dataset or mixed multi-datasets. The larger the indicator is, the better the generalization performance is. When the indicator value is close to 1, it indicates that the performance of the model on unseen datasets is equal to that on the seen datasets, indicating excellent model generalization.

Dosimetric analysis

Dosimetric accuracy of sCT images were evaluated by clinical rectal cancer treatment planning. A dose of 5000 cGy was prescribed for the primary tumor target and the photon plan was designed for each test data using real CT images (TPS, UIH). Then the segmentation and plan of real CT image were copied to the sCT image. The



Fig. 3 Performance of different models on the seen and unseen datasets

dose distribution of the plan generated on real CT was recalculated on the sCT to investigate the gap between them. The dose matrix has a resolution of $3 \times 3 \times 3$ mm³ and covers the main region of interest (ROI).

Results

Image comparison

The results of the two samples are shown in Fig. 3. The first column shows the input MR image and its highlighted region. The second to the sixth columns show the real CT images, the prediction results of CycleGAN, CycleGAN with CL loss, RTGAN (ResTransformer Generative Adversarial Network), 2.5D RTGAN with CL loss and 2.5D CRGAN with CL loss. The first sample was selected from validation sets of PKPU that can be considered as the seen data since the train set included some other data from PKPU. The second sample was picked from public datasets regarded as unseen data for this whole center data were taken as validation datasets.

SCT generated by each model on the seen data were accepted in most areas, with good contrast between the bone and their surrounding soft tissues in the first sample. However, there were some mismatches appearing in the local bone region in sCT, as shown in Fig. 4b3) and





Fig. 4 sCT images generated by different model. The first and third rows show real MR (a1, c1), real CT (a2, c2) and the sCT images generated by CycleGAN (a3, c3), CycleGAN with CL loss (a4, c4), RTGAN (a5, c5), 2.5D RTGAN with CL loss (a6, c6) and our model (a7, c7). The second and fourth rows highlight the ROIs outlined by the yellow box on each corresponding image

 Table 2
 Comparison of MAE, SSIM and SNRpeak of sCT

 produced by different methods in the whole pelvic region

| 1 / | | | 9 | |
|-------------------------|--------|------------------|---------------|-------|
| Method | MAE↓ | SNRpeak ↑ | SSIM ↑ | GP↑ |
| CycleGAN | 47.129 | 25.167 | 0.978 | 0.545 |
| CycleGAN (w/ CL loss) | 43.753 | 26.211 | 0.985 | 0.645 |
| RTGAN | 43.704 | 25.344 | 0.982 | 0.66 |
| 2.5D RTGAN (w/ CL loss) | 42.557 | 26.268 | 0.986 | 0.690 |
| 2.5D CRGAN (w/ CL loss) | 42.344 | 26.979 | 0.992 | 0.911 |

b5). By adding the CL loss to the model, the edge of bone can be improved in Fig. 4b4) and b6). For the second sample in the unseen dataset, our proposed model (Fig. 4(d7)) presented more accurate shape and values of bone than that of other models (Fig. 4(d3) ~ Fig. 4(d6)) by adding the consistency regularization.

To evaluate the accuracy of sCT images generated for treatment planning, we calculated the MAE, SSIM and SNRpeak for the entire outer contour of each patient in the test set under each model, as shown in Table 2. As can be seen from Table 2, adding the Transformer module or using CL loss could improve the result of sCT. The performance of sCT could be further enhanced by combining both of them into CycleGAN. However, the GP values of the above models were relatively low, and the generalization performance of the model was greatly enhanced after using the CL loss.

The training time of our model with the training dataset is 6 days, and the inference time is 3s per image due to the sliding window strategy on a Nvidia 3090. As a comparison, CycleGAN used 5 days of training time and 2.4 s of inference time with the same equipment.

Model generalization

The Fig. 4 and Table 2 have initially shown us the generalization of different models in section "Image comparison". Figure 3 shows the MAE, SNRpeak and SSIM of different models on the seen and unseen datasets. There was big gap in performance between seen and unseen datasets before adopting consistency regularization. This gap could be well compensated after using consistency regularization in our proposed model.

CL loss

Table 3 shows the effect of CL loss addition on MAE of different models. We reported the MAE of the main organs in abdomen, including bladder, rectum, and femur heads. The performance of sCT was enhanced in both bone and soft tissue regions by the introduction of CL loss. The result of main organs was shown in Table 3. In addition, there was a greater decrease in the MAE in bone region, which is consistent with the phenomenon observed in Fig. 4.

Dose comparison and Gamma index

For each patient, a photon plan using Volumetric Modulated Arc Therapy (VMAT) was generated and the DVH was analyzed for target and critical structures. DVH parameters such as mean dose (Dmean), maximum dose

Table 3 Comparison of MAE on the soft tissue and bone region for the sCT generated through different methods

| I | | | 5 | | | |
|--------------|----------|--------------|--------|--------------|---------------|--------------|
| Region | CycleGAN | CycleGAN | RTGAN | 2.5D RTGAN | 2.5D CRGAN | 2.5D CRGAN |
| | | (w/ CL loss) | | (w/ CL loss) | (w/o CL loss) | (w/ CL loss) |
| Bladder | 61.45 | 47.92 | 48.23 | 52.67 | 39.82 | 35.84 |
| Rectum | 59.66 | 47.02 | 45.45 | 42.52 | 42.12 | 40.44 |
| Femur_Head_L | 163.02 | 157.23 | 140.23 | 121.17 | 134.56 | 123.75 |
| Femur_Head_R | 181.73 | 163.39 | 150.34 | 132.60 | 140.28 | 131.89 |

(Dmax), D95%, and D50% were calculated for Planning Target Volume (PTV), Clinical Target Volume (CTV), bladder, left femur head and right femur head. The prescription dose was 50 Gy and 2 Gy/fraction for treatment.

Figure 5 and Table 4 show the comparison of DVH obtained from the radiation planning dose calculation between real CT and sCT generated by 2.5D CRGAN (w/ CL loss). It can be seen that most of the relative differences in the DVH indicators were less than 1%, indicating that the current sCT can meet the needs of radiotherapy planning.

We also reported the Gamma index [35] as Table 5 to show that the results produced by our results have sufficient Gamma index for clinical use. The gamma indices (3 mm, 3%) have been calculated between three-dimension dose distribution of real CT and those of fake CT generated by proposed methods.

Discussion

In this study, we proposed a new model combining contrastive loss and consistency regularization for pelvic MRI-to-CT synthesis. MR images used in our model were single T2 sequences, as suggested in the previous study [36]. The experiment results in Table 2 reveal our superior performance. Primarily, our model presented Page 8 of 11

excellent generalization and performed better than plain CycleGAN, where MAE decreased from 47.129 to 42.344, SNRpeak improved from 25.167 to 26.979, SSIM increased from 0.978 to 0.992 and GP increased from 0.545 to 0.911. Meanwhile, most of the relative differences in the DVH indicators are less than 1%, which is generally considered clinically acceptable. This level of accuracy suggests that the sCT provides a reliable estimate of the actual radiation exposure received by the patient's tissues.

There are many algorithms have been employed in segmentation tasks to improve the accuracy of segmentation on unseen datasets [23–25], and some of them have obtained similar segmentation results on unseen datasets as on seen datasets [24]. However, it is not required to maintain the value distribution of the original data in the segmentation task. In order to increase the generalization performance of a model, strong data augmentation, such as color augmentation, is often used, which disrupts the value distribution of the input image. For the generation task, since the learning target is the image itself, the value distribution of the image needs to be maintained during the training process, and strong augmentation is difficult to be applied directly. Here we adopted consistency regularization similar to semi-supervised learning



Fig. 5 (a) dose distribution map in CT. (b) dose distribution map in synthetic CT. (c) DVH plot with corresponding PTV and OARs

| DVH indicator | | CT vs. synthetic CT mean relative differences(range) | <i>p</i> - val- ue |
|------------------|--------------------------|--|--------------------------|
| PTV Rectum | | | |
| | D _{mean} (100%) | 0.848(-1.880 to 1.970) | 0.208 |
| | D _{max} (100%) | 0.119(-1.130 to 2.161) | 0.866 |
| | D ₉₅ (100%) | 1.255(0.0200 to 2.492) | 0.972 |
| | D ₅₀ (100%) | 0.945(-0.890 to 1.956) | 0.155 |
| CTV Rectum | | | |
| | D _{mean} (100%) | 0.861(-1.550 to 1.799) | 0.486 |
| | D _{max} (100%) | 0.116(-1.130 to 2.288) | 0.871 |
| | D ₉₅ (100%) | 1.580(0.039 to 2.854) | 0.882 |
| | D ₅₀ (100%) | 0.918(-0.860 to 1.871) | 0.185 |
| Bladder | | | |
| | D _{mean} (100%) | 0.830(-1.520 to 4.751) | 0.839 |
| | D _{max} (100%) | 1.149(-0.360 to 2.439) | 0.236 |
| | D ₉₅ (100%) | 0.677(-3.060 to 1.687) | 0.922 |
| | D ₅₀ (100%) | 0.926(-1.680 to 3.295) | 0.874 |
| Left Femur Head | | | |
| | D _{mean} (100%) | -0.070(-1.110 to 0.871) | 0.340 |
| | D _{max} (100%) | 0.016(-1.610 to 1.626) | 0.334 |
| | D ₉₅ (100%) | 0.260(-1.890 to 1.250) | 0.628 |
| | D ₅₀ (100%) | 0.800(-1.140 to 1.180) | 0.330 |
| Right Femur Head | | | |
| | D _{mean} (100%) | -0.090(-0.750 to 0.530) | 0.980 |
| | D _{max} (100%) | -0.180(-3.400 to 1.395) | 0.968 |
| | D ₉₅ (100%) | -0.180(-1.730 to 3.600) | 0.990 |
| | D ₅₀ (100%) | -0.250(-1.350 to 1.120) | 0.958 |

Table 5 Relative dose differences between sCT and CT plans.>0.05: not significant, paired two tailed t-test

| Region | CycleGAN | CycleGAN (w/CLloss) | RTGAN | 2.5DRT- GAN (w/ CLIoss) | 2.5DCRGAN (w/ CL loss) |
|---------------------|----------|------------------------|-------|----------------------------------|---------------------------|
| External | 0.893 | 0.897 | 0.903 | 0.905 | 0.907 |
| Bladder | 0.902 | 0.852 | 0.848 | 0.903 | 0.951 |
| Femur_ Heads | 0.903 | 0.915 | 0.922 | 0.874 | 0.924 |
| CTV_ Rec- tum | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| PTV_ Rec- tum | 0.997 | 0.996 | 0.996 | 0.995 | 0.999 |

[26]. Weak augmentation was used to maintain the value distribution of MR data during the training of CRGAN and Strong augmentation was used to improve the generalization performance of the model. Consistency regularization was used to ensure that MR images undergoing strong augmentation and weak augmentation generate similar sCT. The generalization of 2.5D RTGAN with CL loss was poor with a GI around 0.7 and it can be

improved considerably to 0.91 in our model using consistency regularization. This can also be seen from the Fig. 4 that the performance of our model on the unseen dataset was close to the performance on the seen dataset. These demonstrate the effectiveness of using strong data augmentation as well as consistency regularization.

Contrastive learning loss has been shown to be effective in the generation task [16–18]. In this study, the semantic relation was introduced into contrastive learning and the hard negative mining strategy was explored based on semantic relation. The semantic relation can enhance the structural consistency of MR and corresponding sCT image blocks. The shape of bone of sCT was significantly improved by adding contrastive learning loss, as shown in Fig. 4. It can also be seen from Table 3 that using contrastive learning can effectively improve the results of sCT, and the improvement is mainly concentrated in the bone region. These show that contrastive learning can enhance the structural consistency of MR and sCT and improve the results of sCT.

In this study, we also embedded the Transformer block into the generator to improve the performance of the model. The Transformer module was placed at the last layer of the encoder to extract deep features more efficiently. Compared with the original CycleGAN, all the performance was improved after adding the Transformer module, indicating the superiority of the Transformer in extracting deep features, which is consistent with the previous studies [33].

The existing clinical workflow can be improved and accelerated the efficiency of initial treatment. For TPS manufacturers, providing an initial model that can be used for multiple centers is very important. It can be used to verify whether the entire treatment workflow can go successfully. After training with specific center data, specialized optimization can be carried out for the center to accelerate the implementation of MR-only workflow.

This study has a few limitations. First, the model's generalization depends heavily on whether strong augmentation can simulate the unseen dataset better. However, Clinical data are usually complex and it is challenging to simulate all clinical data through strong augmentation. The registration results of paired training data are also important, more supervise from experienced physicians is required. Therefore, the next step is to improve the generalization of the model with the help of a small number of unseen datasets. More data with standard consistence are also planned to collect in the future for better performance.

MR equipment with higher magnetic field intensity can provide higher signal-to-noise ratio, which means that when converted to pseudo-CT, the image quality will be better and the details will be clearer. High field MR can provide stronger contrast, making the differentiation

 Table 4
 Relative dose differences between sCT and CT plans.

 >0.05: not significant, paired two tailed t-test

between different tissues more prominent when converted to pseudo-CT, which helps in the development of diagnosis and treatment plans. The field of MR data used in our work is 1.5 T, which may limit the image quality and further in the practical clinical workflow. Finally, this study has explored little about the Transformer module and its inherent mechanism.

Conclusion

In this study, we proposed a new model combining contrastive learning loss and consistency regularization loss for multi-center pelvic MRI-to-CT synthesis. The proposed model used a hybrid CNN and Transformer module as a generator. Our model presented excellent generalization in multi-center datasets. With an application in the pelvic region, in which MRI-CT registration is particularly hard, this method is promising for radiotherapy treatment planning and would ease the clinical workflow whilst potentially improving its accuracy.

Abbreviations

| CT | Computed tomography |
|---------|---|
| MR | Magnetic resonance |
| GAN | Generative adversarial network |
| MAE | Mean absolute error |
| SNRpeak | Peak signal-to-noise ratio |
| SSIM | Structural similarity index |
| GP | Generalization Performance |
| DVH | Dose and volume histogram |
| DDCN | Deep convolutional neural networks |
| CGAN | Conditional generative adversarial network |
| CL | Contrastive learning |
| PUPH | Peking University People's Hospital |
| FRFSE | Fast recovery fast spin-echo |
| TPS | Treatment planning system |
| SRC | Semantic relation consistency |
| JSD | Jensen-Shannon Divergence |
| RTGAN | ResTransformer Generative Adversarial Network |
| VMAT | Volumetric Modulated Arc Therapy |
| Dmean | Mean dose |
| Dmax | Maximum dose |
| CNN | Convolutional neural network |
| PTV | Planning Target Volume |
| CTV | Clinical Target Volume |
| | |

Author contributions

Conception and design, acquisition of data: Xianan LiAnalysis and interpretation of data: Lecheng Jia, Fengyu LinMethodology and Software: Ziquan Wei, Hua Li, Weiqi XiongValidation and Investigation: Fan Chai, Tao LiuDrafting the article or critical: Nan Hong, Min ZhangFinal approval of the manuscript: Xianan LiSupervision and Resources: Yi Wang, Wei Zhang.

Funding

This study was funded by Ministry of Industry and Information Technology of the People's Republic of China (grant number TC210H034).

Data availability

The public datasets are available from https://zenodo.org/record/583096.The datasets from PUPH are not publicly available due to privacy of patients. You can contact the corresponding author for the usage of the datasets.

Declarations

Ethical approval and consent to participate

This study followed all dictates of the Declaration of Helsinki and the Ethics Review Board of Peking University People's Hospital, and the hospital approved the research. Written consent to participate was obtained from the patient. The clinical trial number is NCT 030385256, and the registration date is January 31, 2017.

Competing interests

The authors declare no competing interests.

Conflict of interest

The authors declare that they have no conflict of interest.

Received: 13 February 2024 / Accepted: 10 June 2024 Published online: 09 July 2024

References

- Debois M, Oyen R, Maes F, Verswijvel G, Gatti G, Bosmans H, Feron M, Bellon E, Kutcher G, Van Poppel H, Vanuytse L. The contribution of magnetic resonance imaging to the three-dimensional treatment planning of localized prostate cancer. Int J Radiat Oncol Biol Phys. 1999;45:857–65.
- Tenhunen M, Korhonen J, Kapanen M, Seppälä T, Koivula L, Collan J, Saarilahti K, Visapää H. MRI-only based radiation therapy of prostate cancer: workflow and early clinical experience. Acta Oncol. 2018;57:902–7.
- Kapanen M, Collan J, Beule A, Seppälä T, Saarilahti K, Tenhunen M. Commissioning of MRI-only based treatment planning procedure for external beam radiotherapy of prostate. Magn Reson Med. 2013;70:127–35.
- Pollard JM, Wen Z, Sadagopan R, Wang J, Ibbott GS. The future of imageguided radiotherapy will be MR guided. Br J Radiol. 2017;90:20160667.
- Han X. MR-based synthetic CT generation using a deep convolutional neural network method. Med Phys. 2017;44(4):1408–19. https://doi.org/10.1002/ mp.12155.
- Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio A. 2014. Generative adversarial networks. arXiv.org, arXiv:1406.2661.
- Mirza M, Osindero S. Conditional generative adversarial nets. arXiv org. 2014. arXiv:1411.1784.
- Peng Y, Chen S, Qin A, Chen M, Qi Z. 2020. Magnetic resonance-based synthetic computed tomography images generated using generative adversarial networks for nasopharyngeal carcinoma radiotherapy treatment planning. Radiother Oncol, 150.
- Baydoun A, Xu K, Jin UH, Yang H, Muzic RF. Synthetic ct generation of the pelvis in patients with cervical cancer: a single input approach using generative adversarial network. IEEE Access. 2021;9:17208–21.
- Rezaeijo SM, Chegeni N, Naeini B, Makris F, D., Bakas S. Within-modality synthesis and Novel Radiomic evaluation of Brain MRI scans. Cancers. 2023;15(14):3565. https://doi.org/10.3390/cancers15143565. (PMID: 37509228).
- Rezaeijo SM, Hashemi B, Mofid B, Bakhshandeh M, Mahdavi A, Hashemi MS. The feasibility of a dose painting procedure to treat prostate cancer based on mpMR images and hierarchical clustering. Radiation Oncol (London England). 2021;16(1):182. https://doi.org/10.1186/s13014-021-01906-2. (PMID: 34544468).
- 12. Liu Y, Chen A, Shi H, Huang S, Zheng W, Liu Z, Zhang Q, Yang X. Ct synthesis from mri using multi-cycle gan for head-and-neck radiation therapy. Comput Med Imaging Graph. 2021;91:101953.
- Zhu JY, Park T, Isola P, Efros AA. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE international conference on computer vision (pp. 2223–2232).
- Yang H, Sun J, Carass A, Zhao C, Lee J, Prince JL, Xu Z. Unsupervised MR-to-CT synthesis using structure-constrained CycleGAN. IEEE Trans Med Imaging. 2020;39(12):4249–61. https://doi.org/10.1109/TMI.2020.3015379.
- Jung C, Kwon G, Ye JC. (2022). Exploring patch-wise semantic relation for contrastive learning in image-to-image translation tasks. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 18260–18269).
- Park T, Efros AA, Zhang R, Zhu JY. 2020. Contrastive learning for unpaired image-to-image translation. Computer Vision – ECCV 2020, vol 12354.

- Zheng C, Cham TJ, Cai J. (2021). The spatially-correlative loss for various image translation tasks. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021; pp. 16407–16417.
- Jabbarpour A, Mahdavi SR, Sadr AV, Esmaili G, Shiri I, Zaidi H. Unsupervised pseudo CT generation using heterogenous multicentric CT/MR images and CycleGAN: dosimetric assessment for 3D conformal radiotherapy[J]. Comput Biol Med. 2022;143:105277.
- Brou Boni KND, Klein J, Vanquin L, Wagner A, Lacornerie T, Pasquier D, Reynaert N. MR to CT synthesis with multicenter data in the pelvic area using a conditional generative adversarial network. Phys Med Biol. 2020;65(7):075002. https://doi.org/10.1088/1361-6560/ab7633.
- 21. Vajpayee R, Agrawal V, Krishnamurthi G. Structurally-constrained opticalflow-guided adversarial generation of synthetic CT for MR-only radiotherapy treatment planning [J]. Sci Rep. 2022;12(1).
- Li D, Yang J, Kreis K, Torralba A, Fidler S. Semantic segmentation with generative models: Semi-supervised learning and strong out-of-domain generalization. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021; pp. 8296–8307.
- Kim J, Lee J, Park J, Min D, Sohn K. Pin the memory: learning to generalize semantic segmentation. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022; pp. 4340–4350.
- Zhang L, Wang X, Yang D, Sanford T, Harmon S, Turkbey B, Wood B, Roth H, Myronenko A. Generalizing deep learning for medical image segmentation to unseen domains via Deep Stacked Transformation [J]. IEEE Trans Med Imaging. 2020;39(7):2531–40.
- Riccardo V, Hongseok N, Ozan S, John D, Vittorio M, Silvio S. Generalizing to unseen domains via adversarial data augmentation. In NuerIPS, 2018; pp. 5339–5349.
- Zhang B, Wang Y, Hou W, Wu H, Wang J, Okumura M, Shinozaki T. FlexMatch: boosting semi-supervised learning with Curriculum Pseudo labeling [J], 2021. arXiv.org, arXiv:2110.08263.

- Abuduweili A, Li X, Shi H, Xu CZ, Dou D. Adaptive consistency regularization for semi-supervised transfer learning. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021; pp. 6919–6928.
- Nyholm T, Jonsson J, Sohlin M, Gustafsson C, Kjellén E, Söderström K, Albertsson P, Blomqvist L, Zackrisson B, Olsson L, Gunnlaugsson A. MR and CT data with multiobserver delineations of organs in the pelvic area— part of the Gold Atlas project. Med Phys. 2018;45(3):1295–300.
- Modat M, Ridgway GR, Taylor ZA, Lehmann M, Barnes J, Hawkes D, Fox,., Ourselin S. Fast free-form deformation using graphics processing units. Comput Methods Programs Biomed. 2010;98(3):278–84.
- Kingma DP, Ba J. Adam: a method for stochastic optimization. 2014. Arxiv Preprint Arxiv:14126980.
- He K, Zhang X, Ren S, Sun J. (2015). Delving deep into rectifiers: surpassing human-level performance on imagenet classification. In Proceedings of the IEEE international conference on computer vision, pp. 1026–1034.
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A, Kaiser L, Polosukhin I. Attention is all you need [J]. 2017. arXiv:org, arXiv:1706.03762.
- Khan S, Naseer M, Hayat M, Zamir S, Khan F, Shah M. Transformers in vision: a survey [J]. ACM Computing Surveys (CSUR); 2021.
- Miyato T, Kataoka T, Koyama M, Yoshida Y. Spectral normalization for generative adversarial networks. International Conference on Learning Representations, 2018.
- Low DA, Harms WB, Mutic S, Purdy JA. A technique for the quantitative evaluation of dose distributions. Med Phys. 1998;25(5):656–61. https://doi. org/10.1118/1.598248.
- Bird D, Nix MG, Mccallum H, Teo M, Henry AM. Multicentre, deep learning, synthetic-ct generation for ano-rectal mr-only radiotherapy treatment planning. Radiother Oncol. 2021;156(3):23–8.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.