UNPAIRED MR TO CT SYNTHESIS WITH EXPLICIT STRUCTURAL CONSTRAINED ADVERSARIAL LEARNING

Yunhao Ge[1,3,#], *Dongming Wei*[2,3,#] , *Zhong Xue*[3], *Qian Wang*[2], *Xiang Zhou*[3], *Yiqiang Zhan*[3], *Shu Liao*[3,*]

[1] Robotics Institute of Shanghai Jiao Tong University, Shanghai, China
[2] Institute for Medical Imaging Technology, School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China
[3] Shanghai United Imaging Intelligence Co., Ltd, Shanghai, China

ABSTRACT

In medical imaging such as PET-MR attenuation correction and MRI-guided radiation therapy, synthesizing CT images from MR images plays an important role in obtaining tissue density properties. Recently deep-learning-based image synthesis techniques have attracted much attention because of their superior ability for image mapping and faster speed than traditional models. However, most of the current deep-learning-based synthesis methods require large scales of paired data, which greatly limits their usage as in some situation strictly registered image pair is infeasible to obtain. Efforts have been made to relax such a restriction, and the cycle-consistent adversarial networks (Cycle-GAN) is an example to synthesize medical images with unpaired data for training. In Cycle-GAN, the cycle consistency loss is employed as an indirect structural similarity metric between the input and the synthesized images and often leads to mismatch of anatomical structures in the synthesized results. To overcome this shortcoming, we propose to (1) use the mutual information loss to directly enforce the structural similarity between the input MR and the synthesized CT image and (2) to incorporate the shape consistency information to improve the synthesis result. Experimental results demonstrate that the proposed method can achieve better performance both qualitatively and quantitatively for whole-body MR to CT synthesis with unpaired training images compared to Cycle-GAN.

Index Terms**—** image synthesis, training with unpaired data, mutual information, adversarial learning, cross modality

1. INTRODUCTION

The advantage of CT imaging is that its voxel intensities or Hounsfield values directly reflect tissue densities, which can be used for attenuation correction in PET reconstruction and for simulating radiation doses in radiotherapy. With the new development of PET-MR and MR-guided radiotherapy equipment, CT images are no longer acquired, and synthesizing CT from MR images plays an important role in these settings. In addition to matching-learning-based regression models, recently, deep-learning (DL)-based image synthesis techniques have received much attention, and their effectiveness in image synthesis has been well demonstrated [1, 2]. Compared to traditional methods, DL-based methods can simulate images more accurately and have faster inference speed. However, image synthesis across different modalities remains a challenging task due to three reasons: first, most of the DL-based image synthesis methods require a large number of registered image pairs (e.g., the Pix2pix algorithm [1, 3]), which in some situation it is infeasible to obtain; second, the image appearance between two different image modalities can be significantly different. For instance, the bone regions in CT normally have high intensity and can be easily distinguished from other surrounding soft tissues, however the bone regions in MR normally have much lower contrast compared to the surrounding soft tissues [1, 2, 4]; third, the field of view (FOV) between two different modalities can be different, and some voxels in one modality might not have correspondences in the other modality.

Efforts have been made to resolve the above challenges. For instance, the cycle-consistent adversarial networks (Cycle-GAN) [5] is one of the state-of-the-art image synthesis methods, which does not require the existence of registered images of two image modalities (i.e., unpaired image synthesis). However, the cycle consistency loss or Cycle-GAN loss is an indirect constraint to enforce the structural similarity between the source and synthesized image. Specifically, it only requires that the backward transformed image is similar to the original source image, and there is no explicit constraint enforced on the forward transformed image.

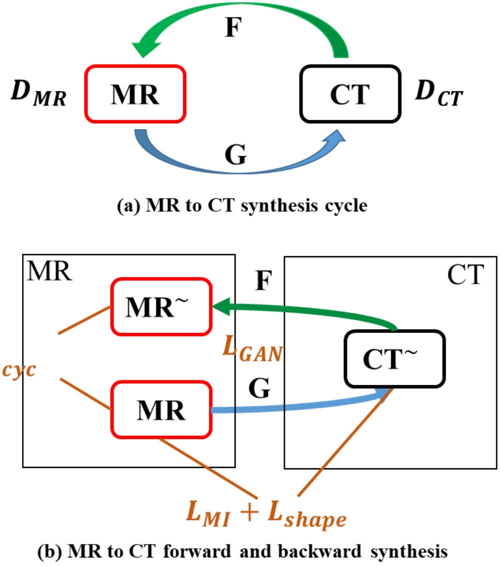
In this paper, we propose a new method to resolve the above challenges. First, a novel mutual information (MI) loss is proposed to directly enforce the structural similarity

# Joint first authors. The first two authors contributed equally to this work. * Corresponding author (email: shu.liao@united-imaging.com).

between the input and synthesized images, which not only improves the representation capability of the network but also boosts the structural consistency between them. Second, we employ a shape consistency (SC) constraint as the second layer of information to further improve the synthesis quality. The proposed method is evaluated on the whole-body MR to CT synthesis problem for automatic PET-MR attenuation correction. Experimental results demonstrate that the proposed method can achieve better MR to CT synthesis results both qualitatively and quantitatively compared with Cycle-GAN.

2. METHOD

The MR to CT synthesis problem for automatic PET-MR attenuation correction is challenging for learning-based techniques because although the MR and CT image pairs of the same patient can be obtained, the patient normally has different articulated local motions at different anatomies. Therefore, it is very hard to perfectly register the MR and CT images even for the same patient (i.e., the unpaired image synthesis problem). Cycle-GAN [5] is one of the state-of-the-art unpaired image synthesis algorithms, and its principle is shown in **Fig. 1(a)**.



(a) MR to CT synthesis cycle

(b) MR to CT forward and backward synthesis

Fig. 1. Schematic illustration of: **(a)** the conventional Cycle-GAN algorithm, and **(b)** the forward and backward synthesis step of the proposed method with MI and SC as the direct structural similarity constraint.

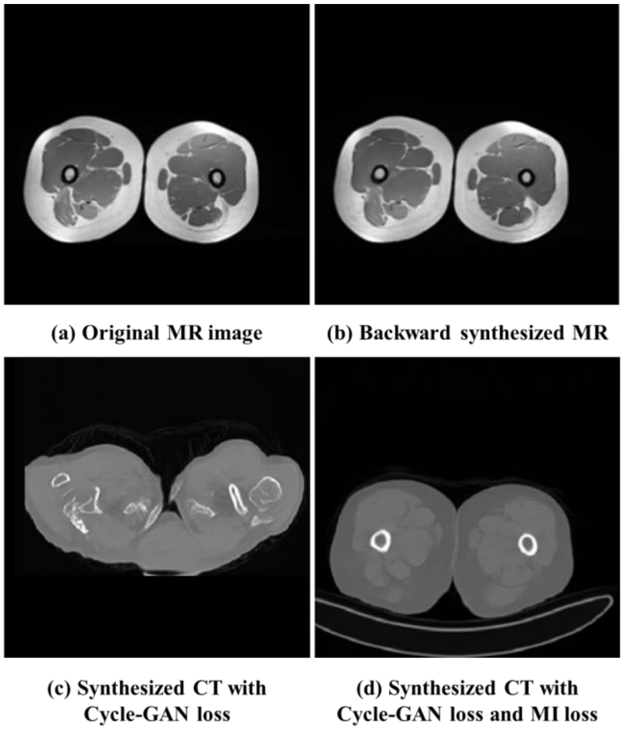
It uses a forward network G to simulate CT from MR, and then uses a backward network F to recover the input. However, Cycle-GAN enforces the structural similarity between the MR and synthesized CT image in an indirect manner through the cycle consistency loss, which may lead to inferior synthesis results. We will analyze this in detail in Section 2.1 and introduce the proposed MI loss during learning to directly enforce the structural similarity constraint in Section 2.2 and the shape consistency constraint to further improve the synthesis results in Section 2.3.

2.1 Cycle consistency loss

The cycle consistency loss is defined by the summation of the similarity between the input MR image I_{MR} and the backward transformed image $F(G(I_{\mathrm{MR}}))$ and the similarity between the input CT image I_{CT} and the backward transformed image $F(G(I_{\mathrm{CT}}))$ as,

$$L_{cyc}(G,F) = \|F(G(I_{MR})) - I_{MR}\|_1 + \|G(F(I_{CT})) - I_{CT}\|_1,$$

thus, the main drawback is that it does not directly enforce the structural similarity between the MR and simulated CT images. Specifically, it is possible that the backward transformed $F(G(I_{\mathrm{MR}}))$ is similar to the original input I_{MR}, but $G(I_{\mathrm{MR}})$ appeared to have strange shapes (see **Fig. 2(c)**).



(a) Original MR image

(b) Backward synthesized MR

(c) Synthesized CT with Cycle-GAN loss

(d) Synthesized CT with Cycle-GAN loss and MI loss

Fig. 2. Synthesized results using the Cycle-GAN loss and MI loss. **(a)** original MR image; **(b)** backward synthesized MR; **(c)** synthesized results using Cycle-GAN loss; and **(d)** synthesized CT with Cycle-GAN loss and MI loss.

2.2. Explicit structural similarity constraint with mutual information

To solve the problem, we propose to explicitly enforce the structural constraint between the input MR image I_{MR} and its synthesized result $G(I_{\mathrm{MR}})$. The main challenge is that during the training stage, we do not have the ground truth $G(I_{\mathrm{MR}})$ to compare under the unpaired setting. We propose to use the MI to enforce this constraint: a good synthesis result should be the CT image of the same patient, which is perfectly aligned with its corresponding input MR image. Thus, MI, one of the most commonly used cross-modality image similarity metrics, is suitable to enforce this constraint between $G(I_{\mathrm{MR}})$ and I_{MR}, and vice versa for $F(I_{\mathrm{CT}})$ and I_{CT}. The schematic illustration of this idea is shown in **Fig. 1 (b)**. The MI loss is defined as,

$$\boldsymbol{L_{MI}} = \sum_{y \epsilon G(I_{MR})} \sum_{x \epsilon I_{MR}} \boldsymbol{p(x,y)}\log\left(\frac{\boldsymbol{p(x,y)}}{\boldsymbol{p(x)p(y)}}\right),$$

where $p(\mathrm{x})$ and $p(\mathrm{y})$ denote the distribution of I_{MR} and $G(I_{\mathrm{MR}})$ respectively, and $p(\mathrm{x,y})$ donates the joint distribution of I_{MR} and $G(I_{\mathrm{MR}})$.

The advantage of using the MI loss is illustrated in **Fig. 2(d)** compared to the result of using the conventional Cycle-GAN loss in **Fig. 2(c)**. It can be observed that the resulting image $G(I_{\mathrm{MR}})$ is structurally more consistent compared to the original input MR image.

2.3. Shape consistency loss

The MI loss introduced in Section 2.2 explicitly enforces the structural constraint between the MR and synthesized CT images. However, it can also be observed that the skin surface and overall shape of the synthesized CT image are still not well aligned with the original image. Here, we propose a SC loss as the second layer of information used during the training process to further improve the synthesis result. Specifically, we extract shape information from the input MR image and the synthesized CT image and enforce a SC loss between them. The schematic illustration is given in **Fig. 3**.

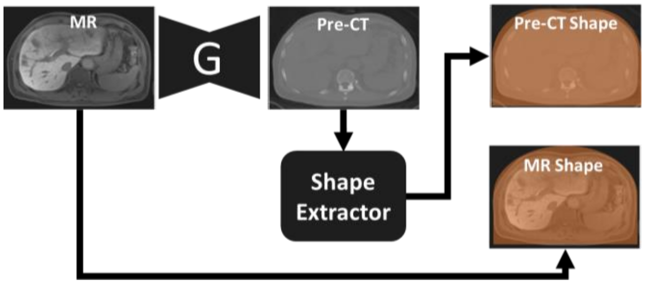


Fig. 3. Schematic illustration of using the SC loss, where the synthesized CT of the generator G will be input into the Shape Extractor network to output shape segmentation. The shape of MR and synthesized CT will be input to compute L1 loss. The images on the right side show overlaying masks on MR and synthesized CT.

As shown in **Fig. 3**, we first annotate the skin surface and body region in both the MR and CT images of each patient in the training data and train a 2D U-Net [6], which extracts the skin surface and body region from the CT image with the training data. Then, during the training process, the shapes from the synthesized CT images are extracted by the trained CNN and compared with the ground truth annotation in the original MR image by L1 loss. Therefore, the forward network generator *G* not only learns to fool the discriminator but also considers the SC loss.

Finally, we adopt the Resnet [7] as the generator's network architecture with 9 residual blocks. For the discriminator, we adopt the Patch-GAN [3] network, which is able to classify whether a local patch is real or fake and has fewer parameter to train compared to conventional convolutional neural networks. The final loss for the forward generator is:

$$\boldsymbol{L} = \boldsymbol{L_{cyc}} + \boldsymbol{L_{shape}} + \boldsymbol{L_{MI}} + \boldsymbol{L_{GAN}}.$$

3. EXPERIMENTAL RESULTS

Fifty patients with Dixon MR sequence and CT scans are used in this study. The MR and CT images of each patient are whole-body scans obtained at different time points and with different motions and FOVs. The Dixon sequence contains the water, fat, in-phase and out-phase channels MR images with resolution 0.91 mm × 0.91 mm × 2.00 mm, and the original CT images have resolution 0.98 mm × 0.98 mm × 1.00 mm. The MR image size is 384×384×476. The CT image size is 512×512×938. The in-phase channel is used to synthesize the CT image. Ten-fold cross validation strategy is used. The Cycle-GAN is trained using Adam for 30 epochs at fixed learning rate of 0.0002 with momentum of 0.5.

3.1 Data preprocessing

All MR and CT images are resampled to have the same resolution of 1 mm × 1 mm × 1 mm. MR and CT images are normalized from [0,1000] and [-1100,2100] to [0,1], respectively. Then, we perform rigid registration between the MR and CT images to remove the global motion of whole body. Finally, 2D axial slices are extracted from 3D MR and CT images, and totally 40 subjects including 38080 slices of MR and 37520 slices of CT are selected to form the training set, and the rest 10 subjects including 9520 slices of MR and 9380 slices of CT are for testing.

3.2 Qualitative Evaluation

We first evaluate the proposed method in a qualitative manner. **Fig. 4** shows typical synthesized results using different approaches: the conventional Cycle-GAN, the proposed method with MI loss only, and the proposed method with both MI and the SC loss constraints.

It can be observed from **Fig. 4** that using the conventional Cycle-GAN, poor synthesized results are obtained. For instance, the soft tissues are deformed in an unrealistic manner, and some parts of the bones are disappeared in the synthesized results. The main reason is that the cycle consistency loss only enforces weak and implicit structural similarity constraint between the input MR and synthesized CT images, and this limitation becomes more obvious for the whole-body scans as there are more anatomical variations in those scans.

By using MI to explicitly enforce the structural similarity constraint between the source MR and synthesized CT images, the synthesized results are much better than using Cycle-GAN. For example, the missing bone regions are restored, and most of the soft tissues are aligned with the source MR images. Therefore, the effectiveness of the MI-based explicit structural similarity constraint is reflected.

It can also be observed from **Fig. 4** that by adding the SC loss together with MI, the skin surface and overall shape of the synthesized results are more similar to the source MR images compared to using MI alone, and the small unrealistic soft tissue deformation existed in the results using MI alone is also corrected.

3.3 Quantitative Evaluation

The mean absolute error (MAE) and peak-signal-to-noise ratio (PSNR) metrics are used to quantitatively evaluate the methods.

Table 1. Mean absolute error (MAE) and peak-signal-to-noise ratio (PSNR) for different anatomies.

Anatomies		Pelvic bones	Lungs	Spine	Femur bones	Average
MAE	Cycle-GAN	107.03	108.53	109.4	104.03	107.25
	MI	93.36	96.89	98.99	90.85	95.02
	MI+SC	**78.34**	**80.24**	**84.00**	**76.30**	**79.72**
PSNR	Cycle-GAN	43.22	43.16	43.13	43.34	43.21
	MI	43.82	43.65	43.56	43.93	43.74
	MI+SC	**44.69**	**44.47**	**44.27**	**44.69**	**44.50**

Notice that MAE and PSNR requires that MR and CT images are registered. However, since the MR and CT images are obtained at different time-points with different local anatomical motions and FOVs, it is difficult to perfectly register the MR and CT images. Therefore, we performed adaptive registration on four different anatomical regions between the MR and CT images: Pelvic bones, Lungs, Spine and Femur bones. Specifically, we manually drew a binary mask for each anatomical region and performed registration for regions only within the mask. For each region, we first performed rigid registration and then performed deformable registration [8]. **Table 1** shows the quantitative evaluation results. It can be observed that by using the MI loss (i.e., "MI" in **Table 1**) alone, we can already obtain better synthesis results compared to Cycle-GAN [6]. By using MI loss and SC loss, the result can be further improved.

4. CONCLUSION

We proposed an adversarial learning method for robust whole-body MR to CT image synthesis with unpaired data training. There are two main contributions of our method. First, we directly enforce the structural similarity constraint by using the MI loss, which is shown to be more robust compared to the cycle consistency loss. Second, the SC loss is used during training to provide a second layer of information to improve the robustness of the synthesis process. Our method has been evaluated both qualitatively and quantitatively and compared with the Cycle-GAN image synthesis method. Experimental results showed that our method consistently achieved better synthesis results for synthesizing CT from MR images.

REFERENCES

[1] L. Xiang, Q. Wang, D. Nie, L. Zhang, X. Jin, Y. Qiao, and D. Shen, "Deep embedding convolutional neural network for synthesizing CT image from T1-Weighted MR image," *Med. Image Anal.*, vol. 47, pp. 31–44, 2018.

[2] D. Nie, R. Trullo, J. Lian, L. Wang, C. Petitjean, S. Ruan, Q. Wang, and D. Shen, "Medical Image Synthesis with Deep Convolutional Adversarial Networks," *IEEE Trans. Biomed. Eng.*, (in press), 2018.

[3] P. Isola, Jun-Yan Zhu, Tinghui Zhou and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," in *Computer Vision and Pattern Recognization (CVPR)*, pp.5967-5976, 2017.

[4] X. Cao, J. Yang, Y. Gao, Q. Wang, and D. Shen, "Region-Adaptive Deformable Registration of CT/MRI Pelvic Images via Learning-Based Image Synthesis," *IEEE Trans. Image Process.*, vol. 27, no. 7, pp. 3500–3512, 2018.

[5] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks," in *IEEE International Conference on Computer Vision (ICCV)*, pp. 2223-2232, 2017.

[6] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention (MCCAI)*, pp.1-8, 2015.

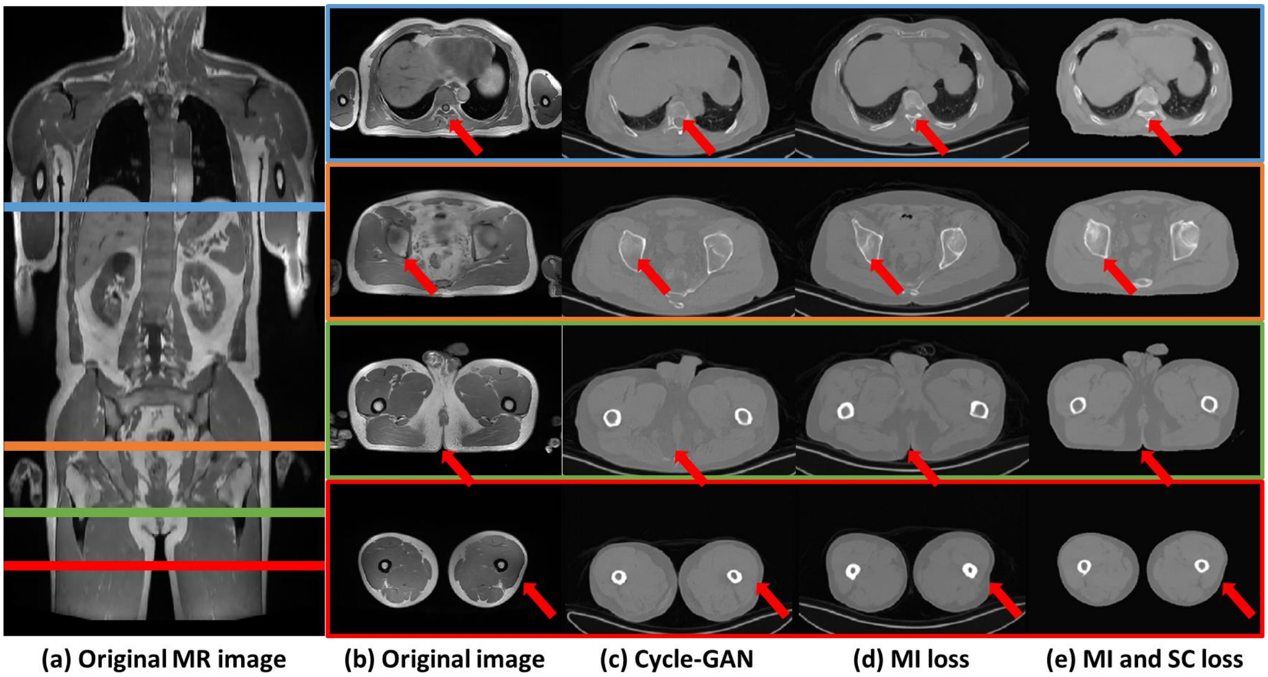


Fig. 4. Performance of different loss functions. It can be seen that MI and SC loss generated realistic CT images with similar shapes of the input MR. Red arrows indicate significant improved regions.

[7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Computer Vision and Pattern Recognization (CVPR)*, pp.770-778, 2016.

[8] B. B. Avants, N. J. Tustison, G. Song, P. A. Cook, A. Klein, and J. C. Gee, "A reproducible evaluation of ANTs similarity metric performance in brain image registration," *Neuroimage*, vol. 54, no. 3, pp. 2033–2044, 2011.