Unpaired whole-body MR to CT synthesis with correlation coefficient constrained adversarial learning

Yunhao Ge [ab], Zhong Xue [b], Tuoyu Cao [c], Shu Liao*[b]

[a] Robotics Institute of Shanghai Jiao Tong University, Shanghai, China; [b] Shanghai United Imaging Intelligence Co., Ltd, Shanghai, China; [c] Shanghai United Imaging Co., Ltd, Shanghai, China

ABSTRACT

MR to CT image synthesis plays an important role in medical image analysis, and its applications included, but not limited to PET-MR attenuation correction and MR only radiation therapy planning. Recently, deep learning-based image synthesis techniques have achieved much success. However, most of the current methods require large scales of paired data from two different modalities, which greatly limits their usage as in some situation paired data is infeasible to obtain. Some efforts have been proposed to relax this constraint such as cycle-consistent adversarial networks (Cycle-GAN). However, the cycle consistency loss is an indirect structural similarity constraint of input and synthesized images, and it can lead to inferior synthesized results. To overcome this challenge, a novel correlation coefficient loss is proposed to directly enforce the structural similarity between MR and synthesized CT image, which can not only improve the representation capability of the network but also guarantee the structure consistency between MR and synthesized CT images. In addition, to overcome the problem of big variance in whole-body mapping, we use the multi-view adversarial learning scheme to combine the complementary information along different directions to provide more robust synthesized results. Experimental results demonstrate that our method can achieve better MR to CT synthesis results both qualitatively and quantitatively with unpaired MR and CT images compared with state-of-the-art methods.

Keywords: image synthesis, unpaired, correlation coefficient constrained, multi-view, whole-body, adversarial learning, MR to CT

1. INTRODUCTION

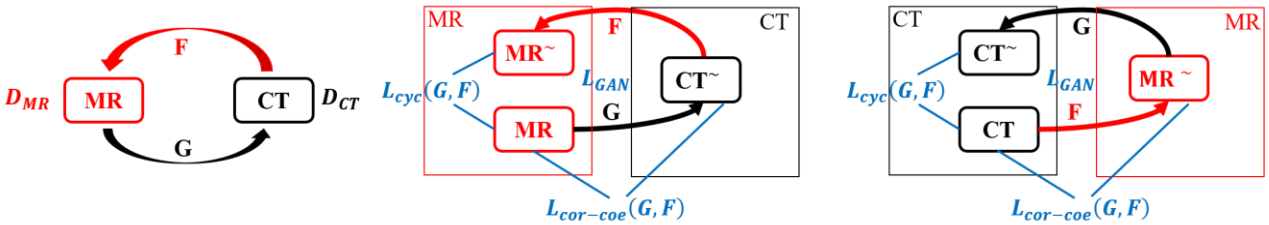
Computed tomography (CT) is critical for various clinical applications, like PET-MR attenuation correction and radiotherapy treatment. However, the exposed radiation during CT acquisition have side effect to patients, while the magnetic resonance image (MRI) is safer without any radiation compared with CT; What is more, some MRI-only radiotherapy treatment and PET-MR attenuation correction only have MR image, getting the new CT image need more cost and not convenient. Therefore, the MR to CT synthesis is highly motivated and very valuable for both scientific research and clinical application. Most of the MR to CT synthesis deep learning methods are supervised learning[1-2], which need a large amount of paired aligned MR image and CT image from the same patient. However, the paired data is extra limited and whole-body registration is difficult and time-consuming for supervised methods that needs paired images to train the model, and the mismatch between paired data may cause errors in synthesized CT image, as different anatomies normally have different local motions. The adaptation of cycle generative adversarial networks (Cycle-GAN)[3] can use the unpaired data to solve some of the supervised learning problems[4]. However, the cycle consistency loss in Cycle GAN is an indirect constraint for the structure consistence between input MR image and synthesized CT, meanwhile, the improvement of cycle consistency loss may damage the mapping ability of generative network. It is hard to explicitly enforce the structural consistency between two modalities in the case of lacking paired training images.

In this paper, a novel correlation coefficient loss is proposed to solve this problem by directly enforce the structural similarity between the input and synthesized images, which not only improves the representation capability of the network but also improve the structure consistency between them. What is more, PET-MR attenuation correction may use fast scan MR image to estimate the CT image, the image resolution in fast scan MR image is low, which easily loose the structure information, makes this MR-to-CT synthesis harder. What's more, the whole-body data, form head to leg, have big variance which require high robust of the synthesized algorithm. Thus, we invented a multi-view adversarial learning to synthesize the corresponding CT image.

2. METHODS

2.1 Correlation coefficient constrain objective function

Cycle-GAN[3] is one of the state-of-the-art unpaired image synthesis algorithms, and its principle is shown in Fig. 1(a). It uses a forward network *G* to simulate CT from MR, and then uses a backward network *F* to recover the input.



(a) MR-to-CT Synthesis cycle (b) Forward Synthesis MR to CT (c) Backward Synthesis CT to MR

Figure 1. (a) illustration of the MR-to-CT synthesis Cycle (b) the forward synthesis from MR to CT and (c) the backward synthesis from CT to MR with correlation coefficient constrain object function.

We first briefly review the loss function of in the Cycle-GAN[3] approach, which is defined in Eq.1.

$$L(G,F,D_{MR},D_{CT}) = L_{GAN}(G,D_{CT},MR,CT) + L_{GAN}(F,D_{MR},CT,MR) + \lambda L_{cyc}(G,F) \quad (1)$$

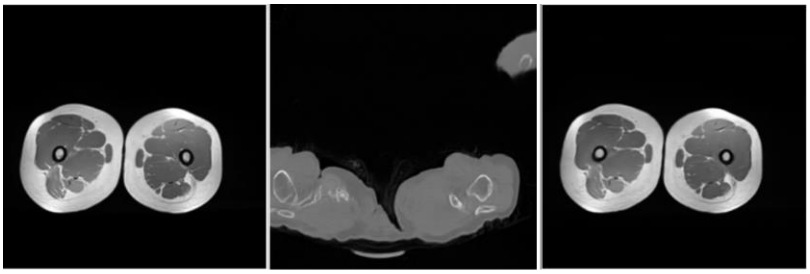
where *G* and *F* denotes the generator networks to map MR to CT and CT to MR, respectively. D_{MR} and D_{CT} denote the discriminator networks in MR and CT domains, respectively. λ is the weight of the cycle consistency loss. $L_{GAN}(G,D_{CT},MR,CT)$ is defined in Eq.2 and $L_{GAN}(F,D_{MR},CT,MR)$ is defined in a similar manner.

$$L_{GAN}(G,D_{CT},MR,CT) = E_{CT\sim P_{data(CT)}}[\log D_{CT}(CT)] + E_{MR\sim P_{data(MR)}}[\log(1-D_{CT}(G(MR)))] \quad (2)$$

The first and second terms in Eq.1 aim to enforce the image appearance similarity for the synthesized results. The third term $L_{cyc}(G,F)$ in Eq.1 is the cycle consistency loss aims to enforce the structural similarity constraint and it is defined in Eq.3

$$L_{cyc}(G,F) = E_{MR\sim P_{data(MR)}}\left[\left\|F(G(MR)) - MR\right\|_1\right] + E_{CT\sim P_{data(CT)}}\left[\left\|G(F(CT)) - CT\right\|_1\right] \quad (3)$$

The main drawback of the cycle consistency loss is it does not directly enforce the structural similarity between the MR and CT images. Specifically, it constrains the mapping from MR to CT with *G* followed by the backward mapping from CT to MR with *F* must be similar to the original input, and vice versa. However, it is possible that *F(G(MR))* is very similar to the original input MR while G and F produce strange mapping effects. Fig 2 illustrates this problem. It is observed that *G(MR)* is very unsatisfactory after applying the generator G. However, after applying the backward generator F, the result image *F(G(MR))* is very similar to the original input MR



(a) Original MR image (b) Synthesized CT with only cycle loss (c) Cyclically synthesized MR

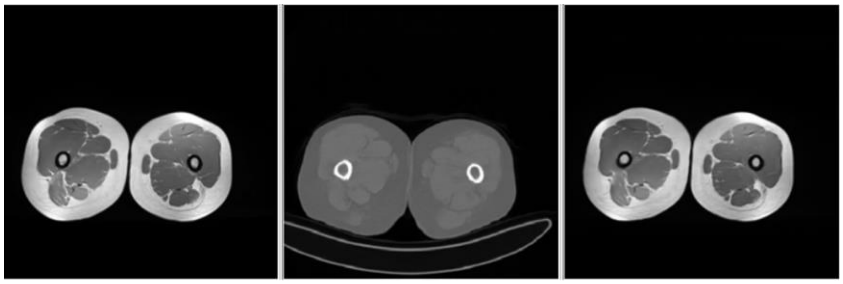
Figure 2. Synthesized results using Cycle-GAN.

In order to resolve this problem, we propose to explicitly enforce the structure constraint between the input MR and its synthesized result *G(MR)*, as is shown in Fig1(b) and (c). The main challenge is during the training phase, we don't have the ground truth *G(MR)* to compare under the unpaired image environment. We propose to use the correlation coefficient to enforce this constraint, and its principle can be understood intuitively in this manner: The perfect synthesis result should be the CT image of the same patient which is perfectly aligned with its corresponding input MR image. Thus, the correlation coefficient, one of the most commonly used multi-modality image registration metrics, is very suitable to enforce this constraint between *G(MR)* and MR, and vice versa for *F(CT)* and CT. It is defined in Eq.4.

$$L_{cor-coe}(G,F)=\frac{Cov(G(MR),MR)}{\sigma_{G(MR)}\sigma_{MR}}+\frac{Cov(F(CT),CT)}{\sigma_{F(CT)}\sigma_{CT}} \tag{4}$$

where *Cov* denotes the covariance. σ denote the variance, and the original cycle consistency loss is replaced with the correlation coefficient loss in Eq.1. The final objective function is shown in Eq.5.

$$L(G,F,D_{MR},D_{CT})=L_{GAN}(G,D_{CT},MR,CT)+L_{GAN}(F,D_{MR},CT,MR)+\lambda L_{cyc}(G,F)+\beta L_{cor-coe}(G,F) \tag{5}$$



(a) Original MR image (b) Synthesized CT with correlation coefficient loss (c) Cyclically synthesized MR

Figure 3. Synthesized results with correlation coefficient loss.

The advantage of using the correlation coefficient loss is illustrated in Fig 3. It can be observed that the resulting image *G(MR)* is structurally more consistent compared to the original input MR image.

2.2 Architecture of generator and discriminator

There are two networks in the Cycle-Consistent Adversarial Networks, the generator and discriminator. The generator is crucial for the quality of synthesized image. In this paper, we adopt the Resnet[5] as the generator's network architecture with 9 residual blocks, which is capable of learning the sophisticated mapping between MR and CT, as well as training not too hard. The architecture is proposed by Jomson.J[5] and used in[3], which are 2D fully convolutional networks, containing two stride convolution layers, nine residual blocks[6] and two fractionally stride convolutional layers. The generator can satisfy any size input image. For the discriminator, we adopt the PatchGAN[3] network, which is capable to classify whether a local patch is real or fake and has fewer parameter to train compared to conventional convolutional neural networks.

2.3 Multi-view adversarial learning

A straightforward solution to synthesize 3D CT volume from MR volume is to synthesize the 2D slices along one direction (e.g., axial direction) and then stack them together. However, such approach has severe drawback as it only considers the information along one direction, and it may produce strange structures and artifacts along the other two directions. Another solution is to directly operate on 3D images, but the computational burden is extremely high.

To address these issues, we propose a multi-view learning strategy illustrated in Fig.4, where synthesized images are obtained along three different directions (i.e., axial, coronal, and sagittal), then we combine the synthesized images from three directions to obtain the final synthesis result. After trying different fusion method, i.e., average, maximum, minimum, normalized-maximum, we adopt the average fusion scheme to obtain the final result.

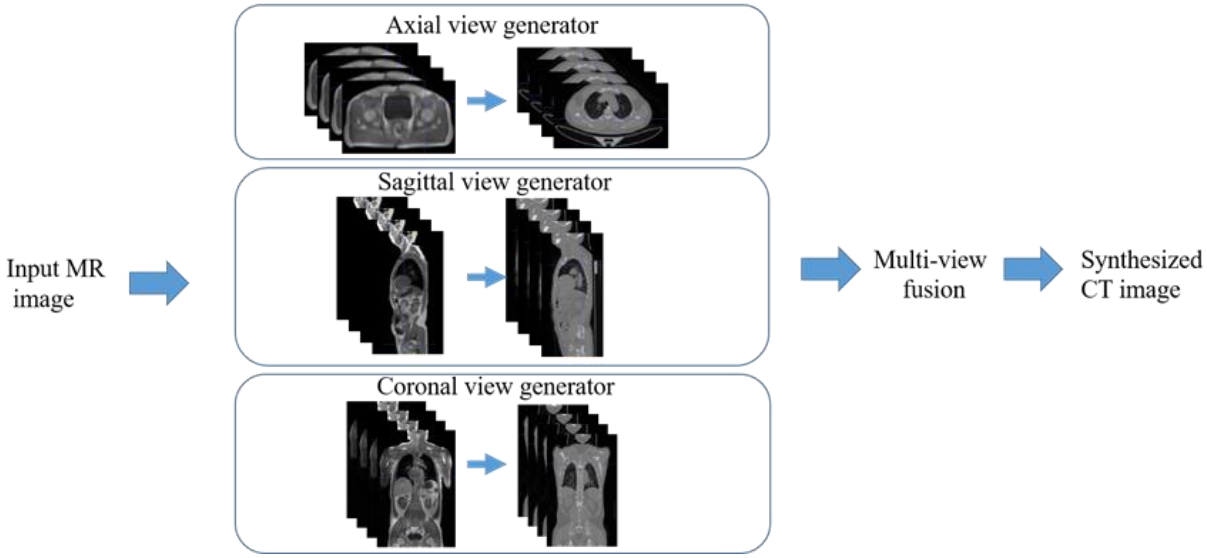


Figure 4. Illustration of our multi-view adversarial learning scheme.

3. EXPERIMENTAL RESULTS

To evaluate our method, we contain two whole-body MR and CT unpaired dataset, high resolution whole-body unpaired dataset (HRWD) and low resolution whole-body unpaired dataset (LRWD). The HRWD contains 50 patients with Dixon MR sequence and CT scans and the LRWD contain 20 patients with Dixon MR sequence and CT scans. In both datasets, The MR and CT images of each patient are scanned at different time points and with different motions and field of views (FOV).

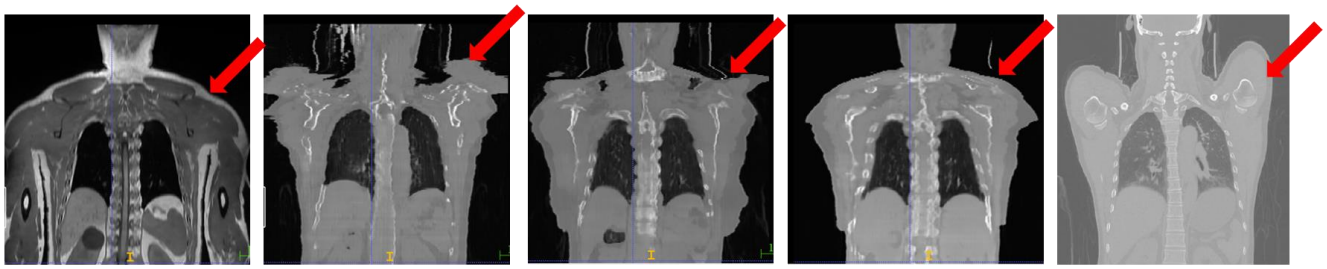
In HRWD, the Dixon sequence contains the water, fat, in-phase and out-phase channels with resolution 0.9766 mm × 0.9766 mm × 2 mm, and the original CT images have resolution 0.91 mm × 0.91 mm × 1 mm, and the CT images are resampled to the same resolution of MR images. Finally, 2D axial slices are extracted 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.

In LRWD, the Dixon sequence contains only the in-phase channels with resolution 2.4 mm × 2.4 mm × 2.4 mm, and the original CT images have resolution 0.9766 mm × 0. 9766 mm × 0. 9766 mm, and the CT images are resampled to the same resolution of MR images. Finally, 2D axial slices are extracted from 3D MR and CT images, and totally 16 subjects including 12760 slices of MR and 12240 slices of CT are selected to form the training set, and the rest 4 subjects including 2870 slices of MR and 2740 slices of CT are for testing.

The in-phase channel is used to synthesize the CT image. In this paper, the 10-fold cross validation strategy is used.

3.1 Impact of the correlation coefficient loss

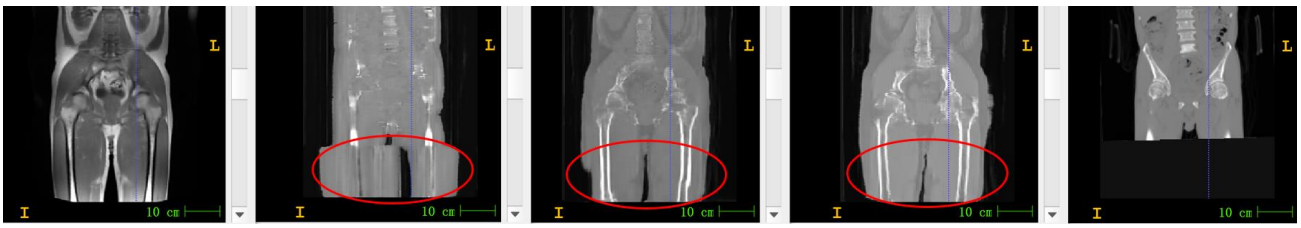
Fig.5 shows the testing result in HRWD with coronal view which shows contribution of using the correlation coefficient loss. Fig.5(a) is an example MR image, and Fig.5(e) is an example CT image, which shows that arms in all MR images are downward while the arms in all CT images are upward. The totally different posture adds tremendous difficulty in mapping task, which requires the synthesized CT images synonymous with the original input MR images with downward arm. However, that kinds of distribution is unreasonable for traditional generator due to the absence of downward CT images in all training data. Fig.4(b) is the synthesis result using Cycle-GAN[3]. It can be observed that the synthesis results are inferior, especially in the arm area, highlighted by the red arrow. Fig.5(c) and Fig.5(d) show the synthesized results with the correlation coefficient loss metric with different weighting parameters. It can be seen that the synthesized arm can be strained from obscure to downward, which satisfy the structure consistence between MR and CT images. When $\beta = 1$, the best results are obtained, and this parameter is fixed in the rest of our experiments.



(a) Original MR image from testing data | (b) Synthesized CT with original loss | (c) Synthesized CT with correlation coefficient loss (β = 0.5) | (d) Synthesized CT with correlation coefficient loss (β = 1) | (e) Original CT image (Unpaired)

Figure 5. Comparison of synthesized CT images with different loss function in HRWD with coronal view.

Fig.6 shows the testing result in LRWD with coronal view which shows contribution of using the correlation coefficient loss. Fig.6(a) is an example MR image, and Fig.6(b) is the synthesis result using Cycle-GAN[3]. It can be observed that the synthesis results are unsatisfactory, especially in the area where the CT image is out of FOV highlighted by the red circle. Fig.6(c) and Fig.6(d) show the synthesized results with the correlation coefficient loss metric with different weighting parameters. It can be shown that systematic improvements are obtained across different anatomies such as bones and fats. When β= 1, the best results are obtained, and this parameter is fixed in the rest of our experiments.



(a) Original MR image | (b) Synthesized CT with original loss | (c) Synthesized CT with correlation coefficient loss(β= 0.5) | (d) Synthesized CT with correlation coefficient loss(β= 1.0) | (e) Original CT image (unpaired)

Figure 6. Comparison of synthesized CT images with different loss function in LRWD with coronal view.

Fig.7 shows the testing result in HRWD with axial view which shows contribution of using the correlation coefficient loss. Column (a) shows the original MR images, column (b) shows the synthesized CT image with original cycle consistent loss and column (c) shows the synthesized CT image with correlation coefficient loss. Row (A) (B) (C) (D) shows different part in whole-body data, the lung part, spine part, pelvic part respectively. In each row, the three images (a), (b) and (c) are linked and the blue cross shows the same point among three images. Because of the different tables between MR and CT devices, it results in the flattened back surface in MR images while a curve back surface in CT images. We can see that the synthesized CT image with original loss in (b) column cannot constrain the back surface, which shows significant gap highlighted in the blue cross. After adding the correlation coefficient loss, the skin surface and overall shape of the synthesized results in (c) column are more similar to the source MR images.

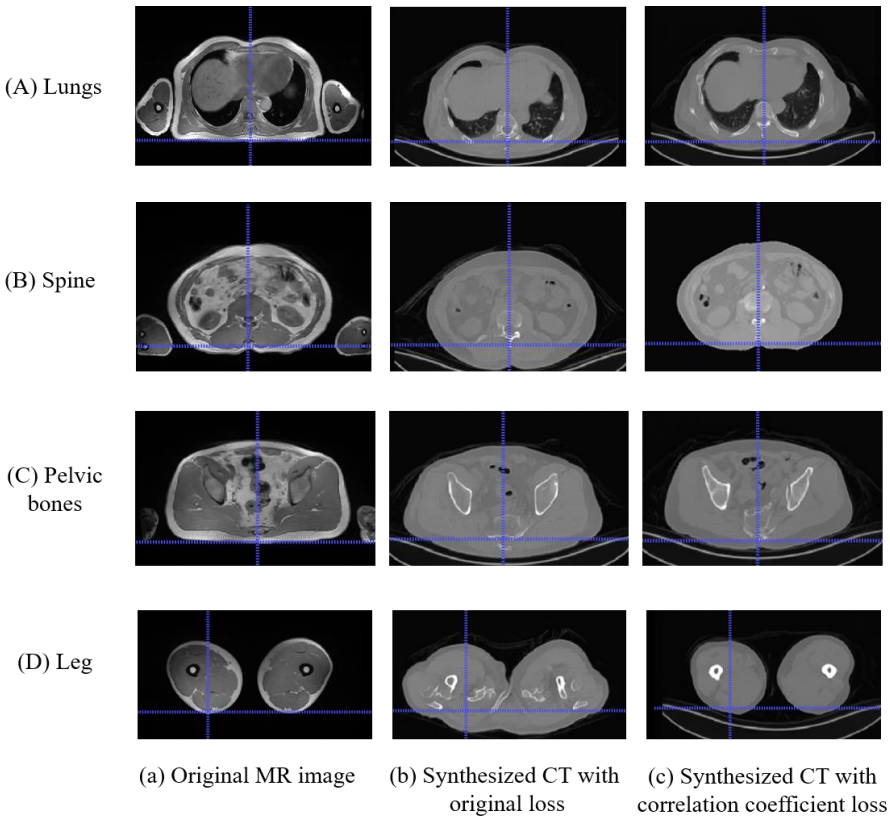
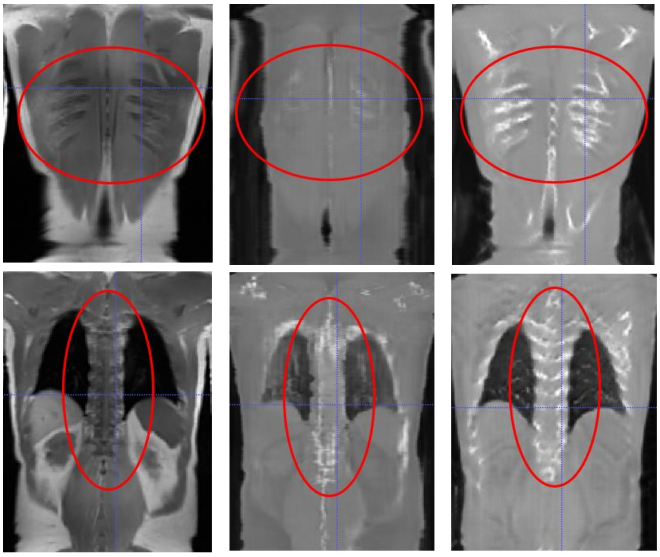


Figure 7. Comparison of synthesized CT images with different loss function in HRWD with axial view. Row (A) to row(D) donates the different part of axial view, lung part, spine part, pelvic bone part and leg part respectively. Column (a) shows the original MR images, column (b) shows the synthesized CT image with original cycle consistent loss and column (c) shows the synthesized CT image with correlation coefficient loss.

3.2 Impact of multi-view adversarial learning

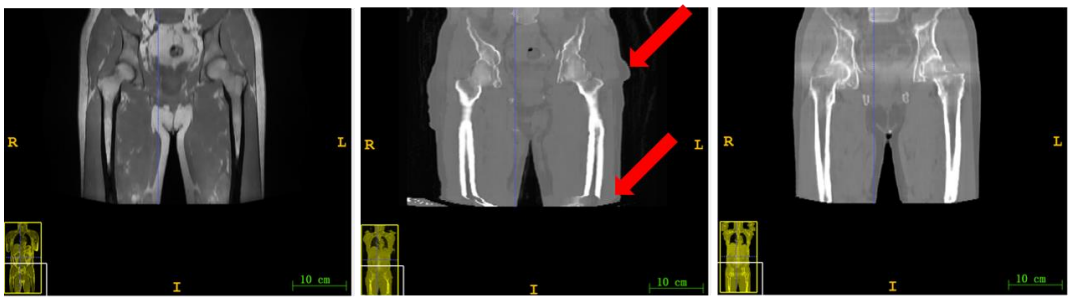
Fig.8 shows the contribution of using the multi-view adversarial learning scheme in LRWD with coronal view. Fig.8(a) shows the original MR image, and Fig.8(b) shows the synthesis results using only the axial direction. It can be observed that the synthesized result is unsatisfactory, especially in the bone and lung areas. Fig.8(c) shows the synthesized results using the multi-view adversarial learning scheme, and it can be observed that systematic improvements are achieved. Regions with significant improvements are highlighted by red circles. Therefore, the advantages of using the multi-view adversarial learning scheme is illustrated.



(a) Original MR image | (b) Synthesized CT with correlation coefficient loss in single view | (c) Synthesized CT with correlation coefficient loss in Multi-view fusion

Figure 8. Advantages of using multi-view adversarial learning to synthesize CT images in LRWD with coronal view.

Fig.9 shows the contribution of using the multi-view adversarial learning scheme in HRWD with coronal view. Fig.9(a) shows the original MR image, and Fig.9(b) shows the synthesis results using only the axial direction. It can be observed that the synthesized result is unsatisfactory, especially in the skin surface and bone. Fig.9(c) shows the synthesized results using the multi-view adversarial learning scheme, and it can be observed that systematic improvements are achieved. Regions with improvements in skin surface and FOV difference between original MR and CT images are highlighted by red arrows. Therefore, the advantages of using the multi-view adversarial learning scheme is illustrated.



(a) Original MR image | (b) Synthesized CT with correlation coefficient loss in single view | (c) Synthesized CT with correlation coefficient loss in Multi-view fusion

Figure 9. Comparison of synthesized CT images with different loss function in HRWD with coronal view.

3.3 Quantitative evaluation

Our method has been also quantitatively evaluated with the mean absolute error (MAE) and peak-signal-to-noise ratio (PSNR) metrics. In order to compute MAE and PSNR, registration between the MR and CT images is required. Since the MR and CT images are obtained at different times with different local anatomical motions and FOVs, it is difficult or even infeasible to perfectly register the MR and CT images. We perform adaptive registration in HRWD on four

different anatomical regions between the MR and CT images: Pelvic bones, Lungs, Spine and Femur bones. Specifically, we manually draw a binary mask for each anatomical region, and perform registration for regions only within the mask. For each region, we first perform rigid registration, and then perform deformable registration[7], and MAE and PSNR are calculated. Table.1 shows the quantitative evaluation results. It can be observed that by using the correlation coefficient loss (i.e., "Single View" in Table.1) alone, we can already obtain better synthesis results compared to Cycle-GAN[3]. By using multi-view adversarial learning, the result can be further improved.

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

Anatomies	**MAE**			**PSNR**		
	Cycle-GAN [3]	Single View	Our method	Cycle-GAN [3]	Single View	Our method
Pelvic bones	107.0375	93.3646	**78.3420**	43.2265	43.8200	**44.6961**
Lungs	108.5330	96.8915	**80.2480**	43.1660	43.6589	**44.4775**
Spine	109.4070	98.9966	**84.0068**	43.1314	43.5656	**44.2787**
Femur bones	104.0375	90.8569	**76.3082**	43.3499	43.9382	**44.6961**
Average	107.2538	95.0274	**79.7263**	43.2185	43.7457	**44.5086**

4. CONCLUSION

We proposed a novel correlation coefficient constrain and multi-view adversarial learning method for robust whole-body MR-to-CT image synthesis with unpaired data. There are two main contributions of our method: First, we directly enforce the structural similarity constraint by using the correlation coefficient loss, which is shown to be more robust compared to the cycle consistency loss. Second, the multi-view synthesis scheme is used to capture complementary information across different directions. Our method has been evaluated both qualitatively and quantitatively, and it is compared with state-of-the-art Cycle-GAN image synthesis method. Experimental results show that our method consistently achieve better synthesis results for different anatomies, which illustrates the effectiveness of our method.

REFERENCES

[1] Nie D., Cao X., Gao Y., Wang L., Shen D. Estimating CT Image from MRI Data Using 3D Fully Convolutional Networks. DLMIA, pp.170-178, (2016)

[2] Han X. MR-based Synthetic CT Generation using a Deep Convolutional Neural Network Method. Medical Physics, 44(4):1408-1419, (2017)

[3] Zhu J Y, Park T, Isola P, et al. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. ICCV pp.2223-2232, (2017)

[4] Wolterink J M, Dinkla A M, Savenije M H F, et al. Deep MR to CT Synthesis Using Unpaired Data. International Workshop on Simulation and Synthesis in Medical Imaging, pp 14-23, (2017)

[5] Johnson J, Alahi A, Li F F. Perceptual Losses for Real-Time Style Transfer and Super-Resolution, ECCV pp.694-711, (2016).

[6] 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)

[7] Reuckert D and et al. Nonrigid Registration Using Free-Form Deformations: Application to Breast MR Images. IEEE TMI, 18(8):712-721, (1999)