Hindawi
BioMed Research International
Volume 2020, Article ID 5193707, 9 pages
https://doi.org/10.1155/2020/5193707

Research Article

Comparison of Supervised and Unsupervised Deep Learning Methods for Medical Image Synthesis between Computed Tomography and Magnetic Resonance Images

Yafen Li,[1,2] **Wen Li**,[1,2] **Jing Xiong**,[1] **Jun Xia**,[3] **and Yaoqin Xie**[1]

[1]*Institute of Biomedical and Health Engineering, Shenzhen Institute of Advanced Technology, Chinese Academy of Science, Shenzhen 518055, China*
[2]*Shenzhen College of Advanced Technology, University of Chinese Academy of Science, Shenzhen 518055, China*
[3]*Department of Radiology, Shenzhen Second People's Hospital, The First Affiliated Hospital of Shenzhen University, Shenzhen 518035, China*

Correspondence should be addressed to Yaoqin Xie; yq.xie@siat.ac.cn

Received 7 August 2020; Revised 8 September 2020; Accepted 23 September 2020; Published 5 November 2020

Academic Editor: Zhiguo Zhou

Cross-modality medical image synthesis between magnetic resonance (MR) images and computed tomography (CT) images has attracted increasing attention in many medical imaging area. Many deep learning methods have been used to generate pseudo-MR/CT images from counterpart modality images. In this study, we used U-Net and Cycle-Consistent Adversarial Networks (CycleGAN), which were typical networks of supervised and unsupervised deep learning methods, respectively, to transform MR/CT images to their counterpart modality. Experimental results show that synthetic images predicted by the proposed U-Net method got lower mean absolute error (MAE), higher structural similarity index (SSIM), and peak signal-to-noise ratio (PSNR) in both directions of CT/MR synthesis, especially in synthetic CT image generation. Though synthetic images by the U-Net method has less contrast information than those by the CycleGAN method, the pixel value profile tendency of the synthetic images by the U-Net method is closer to the ground truth images. This work demonstrated that supervised deep learning method outperforms unsupervised deep learning method in accuracy for medical tasks of MR/CT synthesis.

1. Introduction

Cross-modality medical image synthesis between magnetic resonance (MR) images and computed tomography (CT) images could benefit medical procedures in many ways. As a multiparameter imaging modality, magnetic resonance imaging (MRI) provides a wide range of image contrast mechanisms without ionizing radiation exposure, while CT images outperform MR images in acquisition time and resolution of bone structure. CT is also related with electron density which is critical for PET-CT attenuation correction and radiotherapy treatment planning [1]. Generating synthetic CT (sCT) images from MR images makes it possible to do MR-based attenuation correction in PET-MR system [2–6] and radiation dose calculation in MRI-guided radiotherapy planning [7–9]. Synthesizing MR images from CT images can enlarge the datasets for MR segmentation task and thus improve the accuracy of segmentation [10].

In recent years, there have been many efforts to work on medical image synthesis between MR and CT images. Among all these methods, deep learning method exhibited superior ability of learning a nonlinear mapping from one image domain to another image domain. It can be classified into two categories: supervised and unsupervised deep learning methods. Supervised deep learning methods required paired images for model training. In the MR/CT synthesis task, MR and CT images have to be well-registered at first and then used as inputs and corresponding labels for the neural network model to learn an end-to-end mapping. Nie et al. [11] used three-dimensional paired MR/CT image patches to train a three-layer fully convolutional network for estimating CT images from MR images.

Other researchers [4, 5, 12–15] have trained deeper network for MR-based CT image prediction. However, as for medical image dataset, it is not that easy to get paired MR and CT images. It may take a long time span to collect patients who are scanned by both MR and CT scanners. Registration of certain accuracy between MR and CT images are also necessary to make paired MR-CT dataset.

Unsupervised deep learning methods enabled the possibility of using unpaired images for image-to-image translation [16–20]. It was first proposed for natural image synthesis and now has been implemented by many researchers for medical image synthesis [10, 21–24]. Chartsias et al. [10] demonstrate the application of CycleGAN in synthesizing cardiac MR images from CT images, using MR and CT images of different patients. Nie et al. [21] synthesized MR images from CT images with a deep convolutional adversarial network. Since there are plenty of unpaired medical images, the available datasets could be easily enlarged.

Unlike natural images, accuracy is highly emphasized in medical images. In this paper, we aim to compare the accuracy of supervised and unsupervised learning-based image synthesis methods for pseudo-MR/CT generation tasks. Two typical networks of U-Net [25] and CycleGAN [17] were introduced as representatives of supervised and unsupervised learning methods, respectively. Mean absolute error (MAE), structural similarity index (SSIM), and peak signal-to-noise ratio (PSNR) of the synthetic results were calculated to evaluate their performance quantitatively. More detailed comparisons and discussions about the advantage and disadvantage of these methods are included in Results and Discussion.

2. Materials and Methods

2.1. Neural Network Models. In our experiments of pseudo-MR/CT generation tasks, U-Net and CycleGAN were used as the typical representative network of supervised and unsupervised deep learning methods, respectively.

U-Net has made a great achievement in segmentation tasks [25–29]. The advantage of U-Net is that it could use very few images to make a good performance. In this study, we adapted U-Net to an end-to-end image synthesis task.

The basic architecture of U-Net consists of a contracting part to capture features and a symmetric expanding part to enable precise localization. As shown in Figure 1, we added LeakyReLU [30, 31] as activation operation before convolution operation in the contracting part of the network. Activation operation of LeakyReLU was replaced with ReLU [32] in the expanding part. Batch normalization [33] was introduced to U-Net to enable faster and more stable training. In Figure 1, the number of channels is denoted on top of each of the convolution operation, and the size of feature maps is signed in the parentheses.

In the medical image synthesis task, input image and its corresponding label were fed to the proposed U-Net to train and learn an end-to-end nonlinear mapping between them. Figure 1 illustrated the MR-to-CT synthesis using U-Net architecture, which takes MR images as input and CT images as label to train a synthetic CT generating model. On the contrary, when we use CT images as input and MR images as labels, U-Net could be trained as a synthetic MR-predicting model. The loss function used in the proposed U-Net is

$$L_{\text{U-Net}} = \mathrm{E}_{x,y\sim\hat{P}_{\text{data}}}\left[\|f(x)-y\|_1\right] \tag{1}$$

CycleGAN [17] which is proposed by Zhu et al. could be seen as an updated version of generative adversarial networks (GAN) [16]. GAN methods can learn a nonlinear mapping from input image domain to target image domain by adversarial training. CycleGAN introduced the idea of cycle consistency to general GAN methods. Cycle consistency adds restriction that the generated pseudoimage in target domain should be able to be transformed back to the original input image.

We used the CycleGAN architecture from Zhu et al. [17] for our medical image synthesis task. It takes unpaired MR and CT images as inputs to learn nonlinear mappings between these two image modalities. As illustrated in Figure 2, the CycleGAN architecture has two cycles, forward cycle and backward cycle. The forward cycle consists of three networks: two generative networks of G and F and one discriminator of D_{CT}. The backward cycle uses the same generative networks of F and G and a counterpart discriminator of D_{MR}.

In the forward cycle, network G was used to generate synthetic CT (sCT) from input MR images, while network F generated synthetic MR (sMR) from network G-generated sCT images. Network D_{CT} discriminates whether the generated sCT image is real CT or fake. The backward cycle works just the opposite way. Network F took CT images as input images and generated sMR; then, network G synthesized sCT from the F-generated sMR images. Network D_{MR} was used to distinguish whether the sMR image is real MR or fake.

The adversarial losses of CycleGAN are as follows:

$$\begin{aligned} L_{\text{GAN\_}G\text{\_MR}_{\text{to}}\text{CT}} &= \mathrm{E}_{\text{CT}\sim P_{\text{data}}(\text{CT})}\left[\|\log\left(D_{\text{CT}}(\text{CT})\right)\|_1\right] \\ &\quad + \mathrm{E}_{\text{MR}\sim P_{\text{data}}(\text{MR})}\left[\|\log\left[1-\left(D_{\text{CT}}(G(\text{MR}))\right)\right]\|_1\right], \\ L_{\text{GAN\_}F\text{\_CT}_{\text{to}}\text{MR}} &= \mathrm{E}_{\text{MR}\sim P_{\text{data}}(\text{MR})}\left[\|\log\left(D_{\text{MR}}(\text{MR})\right)\|_1\right] \\ &\quad + \mathrm{E}_{\text{CT}\sim P_{\text{data}}(\text{CT})}\left[\|\log\left[1-\left(D_{\text{MR}}(F(\text{CT}))\right)\right]\|_1\right]. \end{aligned} \tag{2}$$

The cycle-consistency loss consists of forward cycle loss $L_{\text{forward\_cyc}}$ and the backward cycle loss $L_{\text{backward\_cyc}}$. It is represented as follows:

$$\begin{aligned} L_{\text{forward\_cyc}} &= \mathrm{E}_{\text{MR}\sim P_{\text{data}}(\text{MR})}\left[\|\left(F(G(\text{MR}))-\text{MR}\right)\|_1\right], \\ L_{\text{backward\_cyc}} &= \mathrm{E}_{\text{CT}\sim P_{\text{data}}(\text{CT})}\left[\|\left(G(F(\text{CT}))-\text{CT}\right)\|_1\right], \\ L_{\text{Cycle-consistency}} &= L_{\text{forward\_cyc}} + L_{\text{backward\_cyc}}. \end{aligned} \tag{3}$$

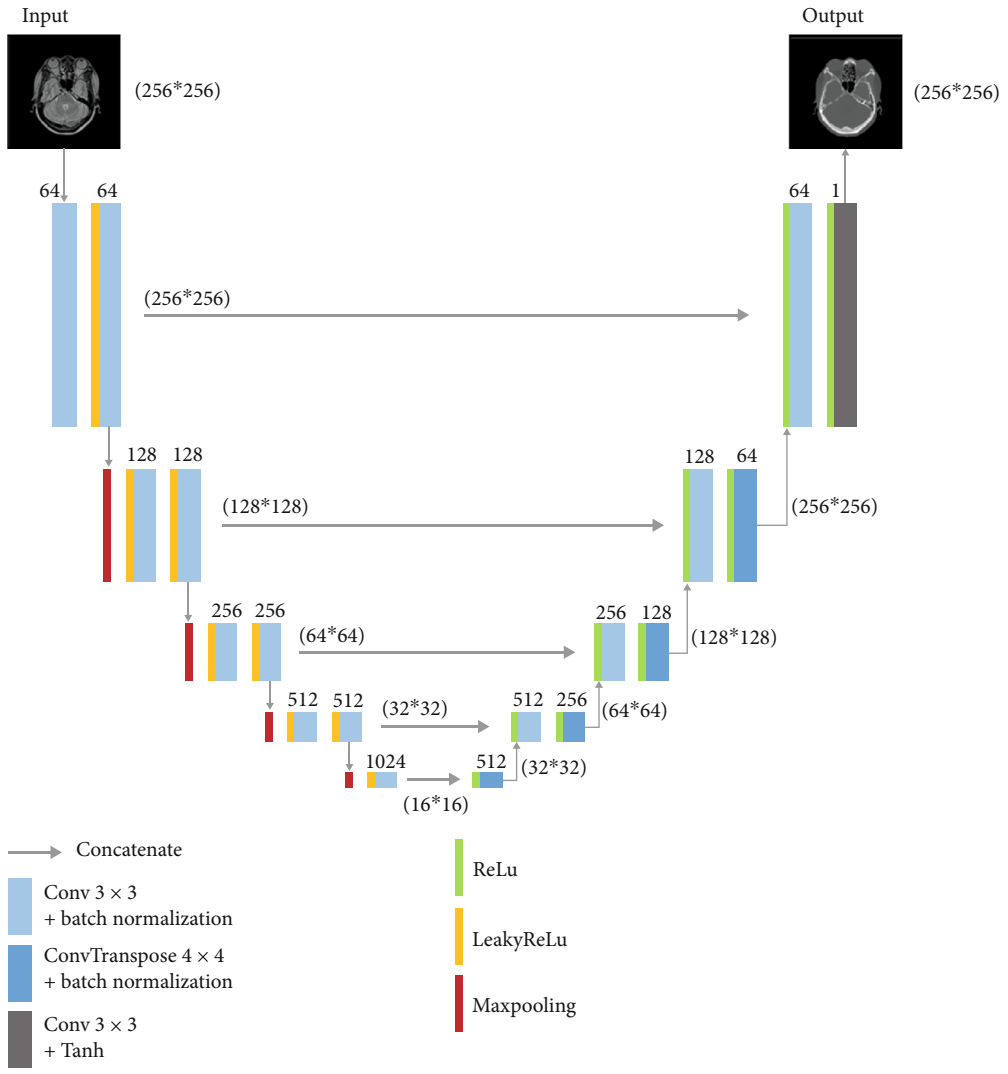

Figure 1: Architecture of proposed U-Net for image synthesis.

Then, we have the full objective as the below equation:

$$L_{\text{CycleGAN}} = L_{\text{GAN\_G\_MR}_{\text{to}}\text{CT}} + L_{\text{GAN\_F\_CT}_{\text{to}}\text{MR}} + \lambda L_{\text{Cycle-consistency}}, \quad (4)$$

where λ is the weight of the objectives of cycle consistency.

2.2. Cross-Modality MR/CT Image Synthesis and Evaluation. We used PyTorch to implement the proposed U-Net and CycleGAN. Both the networks were trained for bidirectional image synthesis, which includes learning a MR-to-CT model for generating synthetic CT images from MR images and a CT-to-MR model for generating synthetic MR images from CT images.

U-Net and CycleGAN used similar parameters for training nonlinear mapping models between MRI/CT images. Adam optimizer was adopted for both the networks. The batch size was set to 1. Both networks were trained for 200 epochs, with fixed learning rate for the first 100 epochs. The learning rate decreased linearly to 0 for the following 100 epochs.

Whole 2D slices of axial medical images with size of $256 * 256$ pixels were used as inputs. During the training process, the images would be padded to $286 * 286$ pixels and then random cropped to $256 * 256$ for data augmentation. While U-Net should utilize paired MR and CT datasets for training nonlinear mapping, CycleGAN can take use of unpaired MR and CT images as inputs for both the forward and backward cycles in training procedure. As for the CycleGAN method, we randomly shuffled the MR image input sequences and CT image input sequences in the paired datasets to make the input MR and CT slices unpaired. The MRI input sequence in unpaired datasets were not the same as that in paired datasets.

Three metrics were used to quantitatively characterize the accuracy of the prediction of synthetic images compared with the ground truth images. The mean absolute error (MAE) measures the discrepancies by voxels. Structural similarity index (SSIM) [34] quantifies the similarities in a whole

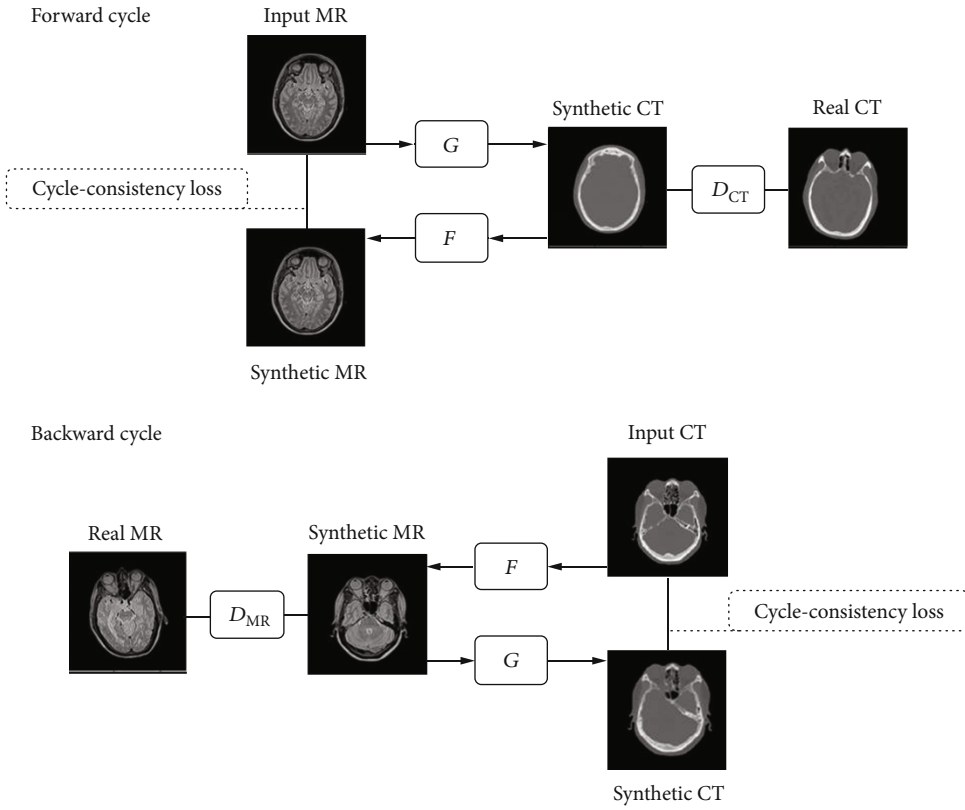

Figure 2: CycleGAN architecture for bidirection synthesis of MR and CT images. The forward cycle generated synthetic CT from input MR by G while F translate the synthetic CT back to the MR image domain. D_{CT} discriminate whether the generated images is real or fake CT. The backward cycle generated synthetic MR from input CT by F while G translate the synthetic MR back to the CT image domain. D_{MR} discriminate whether the generated images is real or fake MR. Two cycle-consistency loss was introduced to capture the intuition that the synthetic image should be translated back to the original image modality.

image scale. Peak signal-noise-ratio (PSNR) assesses the quality of prediction.

These evaluation metrics are expressed as follows:

$$\begin{aligned}
\mathrm{MAE} &= \frac{1}{H * W}\sum_{i=1}^{H}\sum_{j=1}^{W}|X(i,j) - Y(i,j)|,\\
\mathrm{SSIM} &= \left(2\mu_x\mu_y + c_1\right)\left(2\sigma_{xy} + c_2\right)/\left(\mu_x^2 + \mu_y^2 + c_1\right)\\
&\quad \cdot \left(\sigma_x^2 + \sigma_y^2 + c_2\right),\\
c_1 &= (K_1L)^2, c_2 = (K_2L)^2,\\
\mathrm{PSNR} &= 10\ \log_{10}(L/\mathrm{MSE}), \mathrm{MSE} = \frac{1}{H * W}\sum_{i=1}^{H}\sum_{j=1}^{W}(X(i,j) - Y(i,j))^2,
\end{aligned} \tag{5}$$

where H and W are the height and width of the images, respectively. X is the ground truth images, and Y is the predicted synthetic images. μ_x and μ_y are the average values of ground truth images and synthetic images, respectively. σ_x^2 and σ_y^2 are the variance of ground truth images and synthetic images, respectively. σ_{xy} represents the covariance of ground truth images and synthetic images. L denotes the dynamic range of the voxel values. c_1 and c_2 are two variables to stabilize the division with a weak denominator. Here, we take $k_1 = 0.01$ and $k_2 = 0.03$ by default.

2.3. Dataset Preparing. The datasets contain 34 patients. Each patient has both T2-weighted MR images and CT images of the head region. We acquired T2-weighted MR images (TR: 2500 ms, TE: 123 ms, $1 * 1 * 1\,\mathrm{mm}^3$, $256 * 256$) on a 1.5 T Avanto scanner (Siemens). The CT images (120 kV, 330 mA, exposure time: 500 ms, $0.5 * 0.5 * 1$ mm3, $512 * 512$) were acquired on SOMATOM Definition Flash (Siemens).

In this experiment, CT images were resampled to a size of $256 * 256$ ($1 * 1\,\mathrm{mm}^2$) by bicubic interpolation [35] to match the voxel size of MR images. Binary head masks were generated by the Otsu threshold method [36] for MR and CT images to remove unnecessary background information around the head region.

Since the head region is mainly a rigid construction of bone structure, we applied rigid registration to the MR and CT images to make paired MR/CT images for the proposed U-Net. CT images were set as a fixed volume. MR images were set as a moving volume to register with CT images by Elastix toolbox [37]. The paired datasets were randomly shuffled to make an unpaired dataset for CycleGAN.

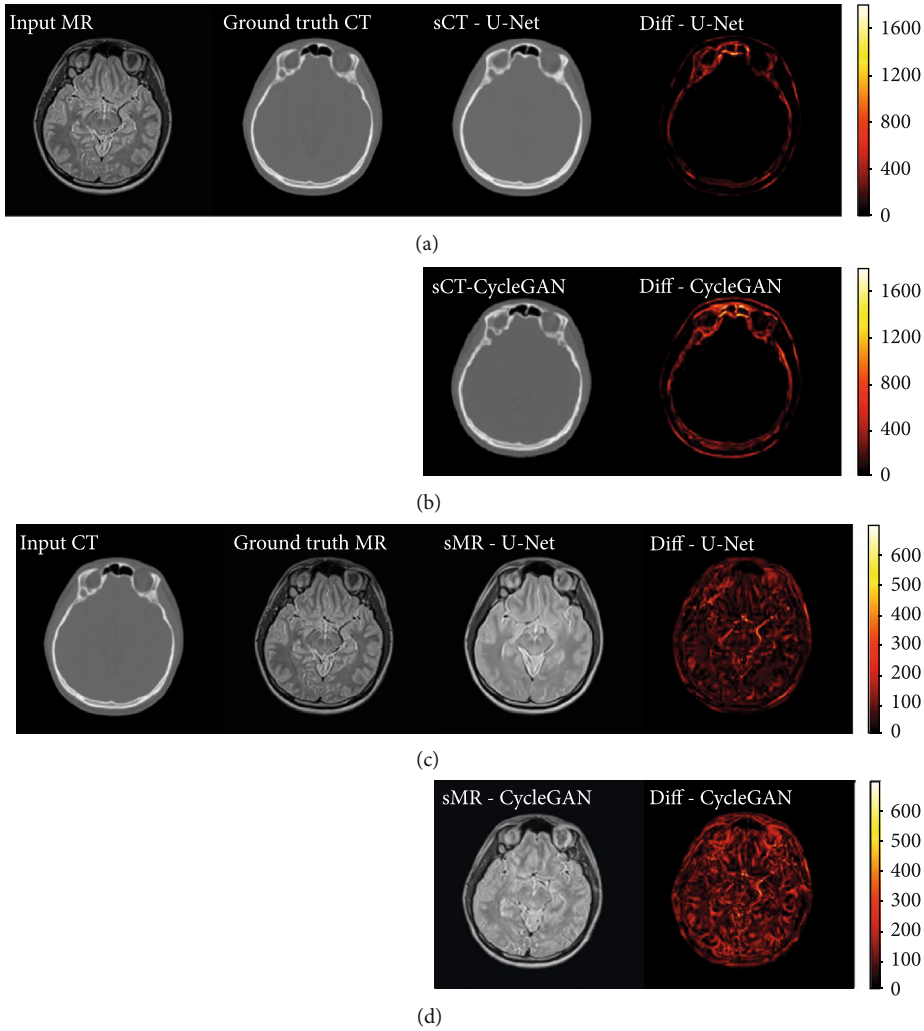

Figure 3: (a–d) From left to right: the 4 columns are input images, ground truth images, synthetic images, and the difference maps. sCT results generated by U-Net (a) and CycleGAN (b), respectively; sMR results by U-Net (c) and CycleGAN (d), respectively.

In our medical image synthesis task, 28 patients with 4063 image pairs were randomly selected for model training. The remaining 6 patients with 846 image pairs were used for evaluation procedure.

3. Results and Discussion

The results of synthetic MR and synthetic CT images generated by U-Net and CycleGAN and their ground truth are showed in Figure 3. The first column is the input images, and the second column is ground truth images. The third column showed the generated synthetic images predicted from input images by the two networks. The difference map between synthetic images and ground truth images was calculated and showed in the fourth column.

The first two rows in Figure 3 are sCT images synthesized by U-Net and CycleGAN, respectively. For the task of synthesizing CT images from MR images, the soft tissue area is translated from high contrast to low contrast. It could be seen from the difference map images that the soft tissue area of synthetic CT images by both networks is well-translated with little error. The translation error mainly occurred in the bone area. Their difference map demonstrates that the sCT by CycleGAN synthesized more error than sCT by U-Net in the bone areas.

The third and fourth rows in Figure 3 are sMR images generated by U-Net and CycleGAN, respectively. It could be seen that sMR by CycleGAN seems more realistic for it has more complex contrast information than sMR by U-Net. However, the difference map images illustrated that the CycleGAN method generated much more error than U-Net does. The abundant image contrast information in sMR by CycleGAN may be false and unnecessary.

In synthesizing CT tasks, the difference between synthetic images and ground truth mainly occurs in the bone area. But in synthesizing MR tasks, the error is evenly distributed in the whole head region. It means synthesizing high contrast images of MR from low contrast image domain of CT is tougher than its reverse synthesizing direction.

To compare the image details, 1D profiles of pixel intensity were also plotted. Figure 4 shows the 1D profiles passing through the short red lines and long blue lines as indicated in

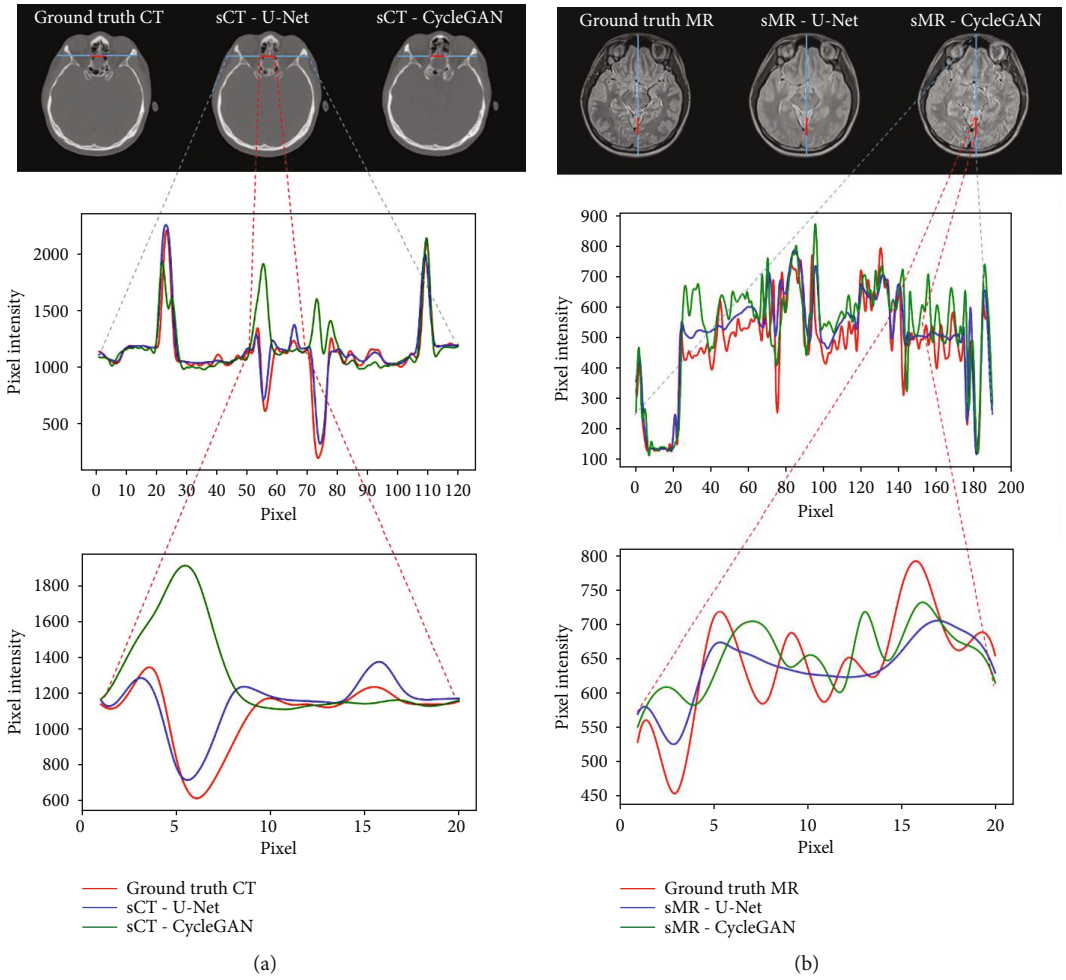

Figure 4: Comparison of 1D profiles of pixel intensity passing through the short red lines and long blue lines as indicated in the images: (a) 1D profile and its close-up marked by the horizontal lines in ground truth CT, U-Net sCT, and CycleGAN sCT images; (b) 1D profile and its close-up marked by the horizontal lines in ground truth MR, U-Net sMR, and CycleGAN sMR images.

the corresponding images in the first row. The red line is overlapped with the blue lines. The 1D profile in the second row of Figure 4 demonstrates pixel intensities of the long blue lines. The 1D profiles in the third row are the pixel intensities of the short red lines of 20 pixels, which shows close-ups of part of the long blue lines' 1D profile.

In the profiles, the red curve indicates pixel intensities of ground truth CT or MR. The blue curve represented for U-Net and the green curve for CycleGAN. It could be clearly seen in Figure 4(a) that the blue curve is close to the red curve, while some of the peaks of the green curve deviated from the red curve to an opposite direction. It means that the tendency of 1D profiles in sCT by U-Net was closer to the ground truth CT, while the CycleGAN method tends to generate fake contrast information in sCT images.

The profile in Figure 4(b) shows that the blue curve vibrated less from the red curve. Some peaks of the green curve deviated more from the red curve. It could be seen in the close-up 1D profile that some peaks of the green curve are biased to the opposite from the red curve, while the tendency of the blue cure seems like a smoothened or flattened red curve. It means that the pixel value of sMR by U-Net was closer to the ground truth but may lack contrast details. The pixel value of sMR by CycleGAN exhibits more deviation from the ground truth along the profile whereas the tendency may be false or exaggerated.

The quantitative metrics have been calculated for comparison. Figure 5 shows the MAE of sCT and sMR for each of the 6 patients in the evaluation datasets and the average result. It is obvious that the U-Net method generated lower MAE either in sCT image generation or sMR image generation for all the patients. This also demonstrates the robust performance of the U-Net method in bidirection MR/CT image translation tasks.

Figures 5(a) and 5(b) show that the deviations of the MAE between the U-Net and CycleGAN method for sMR images of all the 6 patients are not as significant as those for sCT images. In Figure 3, the difference map of sMR indicated that the main predicted errors are evenly distributed in the whole head region, while the main error of sCT mainly occurs mainly in the bone structure. This could be interpreted that generating MR images of high soft tissue contrast

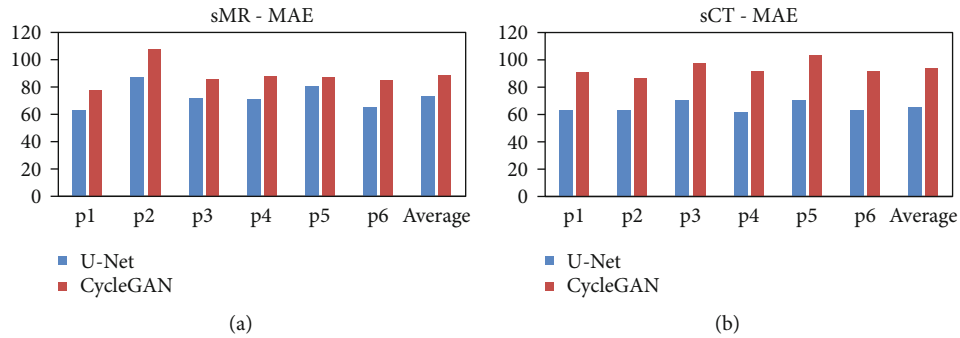

Figure 5: (a) MAE of sMR images for all the 6 patients in test set and their average value. (b) MAE of sCT images generated for all the 6 patients in test set and their average value. Both the blue columns denoted the U-Net method, and the red columns represented the CycleGAN method.

Table 1: Quantitative evaluation results between ground truth CT images and sCT images: MAE, SSIM, and PSNR.

Model	MAE ± SD (HU)	SSIM ± SD	PSNR ± SD (dB)
U-Net	65.36 ± 4.08	0.972 ± 0.004	28.84 ± 0.57
CycleGAN	93.95 ± 5.89	0.955 ± 0.007	26.32 ± 0.55

Table 2: Quantitative evaluation results between ground truth MR images and sMR images: MAE, SSIM, and PSNR.

Model	MAE ± SD (HU)	SSIM ± SD	PSNR ± SD (dB)
U-Net	73.43 ± 9.16	0.946 ± 0.004	32.35 ± 0.78
CycleGAN	88.71 ± 10.04	0.924 ± 0.003	30.79 ± 0.73

from CT images of low soft tissue contrast is much complex than the inverse direction synthesis of generating CT from MR images.

Table 1 shows the overall statistics of three quantitative metrics for sCT by both the U-Net and CycleGAN methods. The SSIM values indicate that the sCT images by both methods have fairly high similarity with the ground truth CT images. The U-Net method outperformed the CycleGAN method with a much lower MAE of 65.36 HU, a higher SSIM of 0.972, and a higher PSNR of 28.84 dB. The average sCT MAE deviation between the two methods is nearly 30 HU.

Table 2 shows the overall statistics of three quantitative metrics for sMR images by the U-Net method and CycleGAN method. The U-Net method outperformed the CycleGAN method with a lower MAE of 73.43 HU, a higher SSIM of 0.946, and a higher PSNR of 32.35 dB.

The qualitative and quantitative results demonstrate that the proposed U-Net, a typical supervised learning method, outperforms CycleGAN, a representative advanced unsupervised learning method, in synthesis accuracy of medical image translation task. Since medical images highly value accuracy for the purpose of disease diagnosing, clinical treatment, and therapeutic effect evaluation, the supervised learning method is more recommended in medical practice.

Nevertheless, the success of supervised learning cannot do without well-registered image pairs. The performance of the trained model also depends on the registration accuracy of the paired images. Unlike natural images, paired medical images are not that easy to get. It would take a long time span to collect enough patients who need to be scanned for both MR and CT images at the same time. It is well-known that big amount of datasets could greatly improve the performance of the deep learning method. Though it outperforms the unsupervised learning method, the limit of dataset volume may constrain the further improvement of the supervised learning method in medical image synthesis tasks.

From the experiments discussed above, the image synthesis by using unsupervised learning methods still has a long way to go for practical application in clinic due to their relatively low accuracy. But still, the unsupervised learning method could benefit when there is lack of paired medical image datasets. The good news is that there are abundant easy-to-obtain retrospective unpaired MR and CT images for the unsupervised learning method to take advantage of. No registration is needed.

Our experiments show that when the same datasets were taken as inputs, the unsupervised learning method got inferior quality in the synthesis accuracy for medical image translation. But nonetheless, if the dataset is large enough, it could be expected that the performance of the unsupervised learning method would be improved to a certain acceptable extent in clinical practice.

4. Conclusions

Cross-modality medical image synthesis between MR and CT images could benefit a lot from the fast growing of deep learning methods. In this paper, we compared different deep learning-based image synthesis methods for pseudo-MR/CT generation, including the unsupervised learning method of CycleGAN and supervised learning methods of the proposed U-Net. Synthetic images produced by the CycleGAN method contain more but fake contrast information in the whole image scale. Though the proposed U-Net method blurred the generated pseudoimages, its pixel value profile tendency is basically close to the ground truth images. The quantitative results also indicate that the U-Net method outperformed the CycleGAN method, especially in synthesizing CT image task.

As accuracy is highly demanded in medical procedures, we recommend the supervised method such as the proposed U-Net in cross-modality medical image synthesis at present clinical practice.

Data Availability

The datasets of MR and CT images used to support the findings in this study are restricted by the Medical Ethics Committee of Shenzhen Second People's Hospital in order to protect patient privacy.

Conflicts of Interest

The authors declare that there is no conflict of interest.

Acknowledgments

This work is supported in part by grants from the National Key Research and Development Program of China (2016YFC0105102), the Leading Talent of Special Support Project in Guangdong (2016TX03R139), the Fundamental Research Program of Shenzhen (JCYJ20170413162458312), the Natural Science Foundation of Guangdong Province (2017B020229002, 2015B020233011, and 2014A030312006), the Beijing Center for Mathematics and Information Interdisciplinary Sciences, and the National Natural Science Foundation of China (61871374).

References

[1] P. Dirix, K. Haustermans, and V. Vandecaveye, "The value of magnetic resonance imaging for radiotherapy planning," *Seminars in Radiation Oncology*, vol. 24, no. 3, pp. 151–159, 2014.

[2] N. Burgos, M. J. Cardoso, K. Thielemans et al., "Attenuation correction synthesis for hybrid PET-MR scanners: application to brain studies," *Ieee Transactions on Medical Imaging*, vol. 33, no. 12, pp. 2332–2341, 2014.

[3] A. Mehranian, H. Arabi, and H. Zaidi, "Vision 20/20: magnetic resonance imaging-guided attenuation correction in PET/MRI: challenges, solutions, and opportunities," *Medical Physics*, vol. 43, no. 3, pp. 1130–1155, 2016.

[4] F. Liu, H. Jang, R. Kijowski, T. Bradshaw, and A. B. McMillan, "Deep learning MR imaging-based attenuation correction for PET/MR imaging," *Radiology*, vol. 286, no. 2, pp. 676–684, 2018.

[5] A. P. Leynes, J. Yang, F. Wiesinger et al., "Zero-echo-time and Dixon deep pseudo-CT (ZeDD CT): direct generation of pseudo-CT images for pelvic PET/MRI attenuation correction using deep convolutional neural networks with multiparametric MRI," *Journal of Nuclear Medicine*, vol. 59, no. 5, pp. 852–858, 2018.

[6] Y. Wu, W. Yang, L. Lu et al., "Prediction of CT substitutes from MR images based on local diffeomorphic mapping for brain PET attenuation correction," *Journal of Nuclear Medicine*, vol. 57, no. 10, pp. 1635–1641, 2016.

[7] M. A. Schmidt and G. S. Payne, "Radiotherapy planning using MRI," *Physics in Medicine and Biology*, vol. 60, no. 22, pp. R323–R361, 2015.

[8] J. A. Dowling, J. Sun, P. Pichler et al., "Automatic substitute computed tomography generation and contouring for magnetic resonance imaging (MRI)-alone external beam radiation therapy from standard MRI sequences," *International Journal of Radiation Oncology • Biology • Physics*, vol. 93, no. 5, pp. 1144–1153, 2015.

[9] A. M. Dinkla, J. M. Wolterink, M. Maspero et al., "MR-only brain radiation therapy: dosimetric evaluation of synthetic CTs generated by a dilated convolutional neural network," *International Journal of Radiation Oncology • Biology • Physics*, vol. 102, no. 4, pp. 801–812, 2018.

[10] A. Chartsias, T. Joyce, R. Dharmakumar, and S. A. Tsaftaris, "Adversarial image synthesis for unpaired multi-modal cardiac data," Springer International Publishing, Cham, 2017.

[11] D. Nie, X. H. Cao, Y. Z. Gao, L. Wang, and D. G. Shen, "Estimating CT image from MRI data using 3D fully convolutional networks," in *Deep Learning and Data Labeling for Medical Applications*, vol. 10008, pp. 170–178, 2016.

[12] X. Han, "MR-based synthetic CT generation using a deep convolutional neural network method," *Medical Physics*, vol. 44, no. 4, pp. 1408–1419, 2017.

[13] F. Liu, P. Yadav, A. M. Baschnagel, and A. B. McMillan, "MR-based treatment planning in radiation therapy using a deep learning approach," *Journal of Applied Clinical Medical Physics*, vol. 20, no. 3, pp. 105–114, 2019.

[14] J. Fu, Y. Yang, K. Singhrao et al., "Deep learning approaches using 2D and 3D convolutional neural networks for generating male pelvic synthetic computed tomography from magnetic resonance imaging," *Medical Physics*, vol. 46, no. 9, pp. 3788–3798, 2019.

[15] K. Gong, J. Yang, K. Kim, G. El Fakhri, Y. Seo, and Q. Li, "Attenuation correction for brain PET imaging using deep neural network based on Dixon and ZTE MR images," *Phys Med Biol*, vol. 63, no. 12, p. 125011, 2018.

[16] I. Goodfellow, J. Pouget-Abadie, M. Mirza et al., "Generative adversarial nets," *Advances in Neural Information Processing Systems 27 (Nips 2014)*, vol. 27, 2014.

[17] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *2017 Ieee International Conference on Computer Vision (Iccv)*, pp. 2242–2251, Venice, Italy, 2017.

[18] Z. L. Yi, H. Zhang, P. Tan, and M. L. Gong, "DualGAN: unsupervised dual learning for image-to-image translation," in *2017 Ieee International Conference on Computer Vision (Iccv)*, pp. 2868–2876, Venice, Italy, 2017.

[19] P. Isola, J. Y. Zhu, T. H. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *30th Ieee Conference on Computer Vision and Pattern Recognition (Cvpr 2017)*, pp. 5967–5976, Hawaii, America, 2017.

[20] J. X. Lin, Y. C. Xia, T. Qin, Z. B. Chen, and T. Y. Liu, "Conditional image-to-image translation," in *2018 Ieee/Cvf Conference on Computer Vision and Pattern Recognition (Cvpr)*, pp. 5524–5532, Salt Lake City, America, 2018.

[21] D. Nie, R. Trullo, J. Lian et al., "Medical image synthesis with context-aware generative adversarial networks," *Medical Image Computing and Computer-Assisted Intervention*, pp. 417–425, 2017.

[22] M. Maspero, M. H. F. Savenije, A. M. Dinkla et al., "Dose evaluation of fast synthetic-CT generation using a generative adversarial network for general pelvis MR-only radiotherapy," *Phys Med Biol*, vol. 63, no. 18, p. 185001, 2018.

[23] H. Emami, M. Dong, S. P. Nejad-Davarani, and C. K. Glide-Hurst, "Generating synthetic CTs from magnetic resonance

images using generative adversarial networks," *Medical Physics*, vol. 45, no. 8, pp. 3627–3636, 2018.

[24] P. Costa, A. Galdran, M. I. Meyer et al., "End-to-end adversarial retinal image synthesis," *Ieee Transactions on Medical Imaging*, vol. 37, no. 3, pp. 781–791, 2018.

[25] O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241, Munich, Germany, 2015.

[26] Y. Yang, C. Feng, and R. Wang, "Automatic segmentation model combining U-Net and level set method for medical images," *Expert Systems with Applications*, vol. 153, p. 113419, 2020.

[27] S. Chen, H. Yang, J. Fu et al., "U-Net plus: deep semantic segmentation for esophagus and esophageal cancer in computed tomography images," *Ieee Access*, vol. 7, pp. 82867–82877, 2019.

[28] N. Abraham and N. M. Khan, "A novel focal Tversky loss function with improved attention U -Net for lesion segmentation," in *2019 Ieee 16th International Symposium on Biomedical Imaging (Isbi 2019)*, pp. 683–687, Venice, Italy, 2019.

[29] Y. Y. Zeng, X. Y. Chen, Y. Zhang, L. F. Bai, and J. Han, "Dense-U-Net: densely connected convolutional network for semantic segmentation with a small number of samples," in *Tenth International Conference on Graphics and Image Processing (Icgip 2018)*, vol. 11069, Chengdu, China, 2019.

[30] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical evaluation of rectified activations in convolutional network. ICML deep learning workshop," 2015, http://arxiv.org/abs/1505.00853.

[31] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proceedings of the 30th International Conference on Machine Learning*, Atlanta, Georgia, USA, 2013.

[32] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proceedings of the 27th international conference on machine learning (ICML-10)*, pp. 807–814, Haifa, Israel, 2010.

[33] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *International conference on machine learning*, pp. 448–456, Lille, France, 2015.

[34] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *Ieee Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[35] H. Prashanth, H. Shashidhara, and K. N. Balasubramanya Murthy, "Image scaling comparison using universal image quality index," in *Advances in Computing, Control, & Telecommunication Technologies, 2009. ACT'09. International Conference*, pp. 859–863, Trivandrum, Kerala, India, December 2009.

[36] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[37] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. W. Pluim, "Elastix: a toolbox for intensity-based medical image registration," *Ieee Transactions on Medical Imaging*, vol. 29, no. 1, pp. 196–205, 2010.