Dual Adversarial Learning with Attention Mechanism for Fine-grained Medical Image Synthesis

Dong Nie[1,2], Lei Xiang[3], Qian Wang[3], Dinggang Shen[1]

[1]Department of Radiology and BRIC, University of North Carolina at Chapel Hill, USA
[2]Department of Computer Science, University of North Carolina at Chapel Hill, USA
[3]Med-X Research Institute, School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China

Abstract. Medical imaging plays a critical role in various clinical applications. However, due to multiple considerations such as cost and risk, the acquisition of certain image modalities could be limited. To address this issue, many cross-modality medical image synthesis methods have been proposed. However, the current methods cannot well model the hard-to-synthesis regions (e.g., tumor or lesion regions). To address this issue, we propose a simple but effective strategy, that is, we propose a dual-discriminator (dual-D) adversarial learning system, in which, a global-D is used to make an overall evaluation for the synthetic image, and a local-D is proposed to densely evaluate the local regions of the synthetic image. More importantly, we build a difficult-region-aware attention mechanism which targets at better modeling hard-to-synthesize regions (e.g., tumor or lesion regions) based on the local-D. Experimental results show the robustness and accuracy of our method in synthesizing target images from the corresponding source images. In particular, we evaluate our method on two datasets, i.e., to address the tasks of generating T2 MRI from T1 MRI for the brain tumor images and generating MRI from CT. Our method outperforms the state-of-the-art methods under comparison in all datasets and tasks. And the proposed difficult-region-aware attention mechanism is also proved to be able to help generate more realistic images, especially for the hard-to-synthesize regions.

1 Introduction

The importance of medical imaging for clinical diagnosis, treatment of disease, and medical research has steadily risen over the last decades. Multiple imaging modalities provide complementary information for each other, which is essential for comprehensive assessment of complex diseases in either diagnostic examinations or as part of medical research trials. For example, computed tomography (CT) is often used for dose planning in radiation therapy for cancer patients. However, CT scan will expose patient to radiation and its image cannot provide good contrast in soft tissue. In contrary, magnetic resonance imaging (MRI) provides very good soft tissue contrast and is much safer than CT (as no radiation

2   Dong Nie[1,2], Lei Xiang[3], Qian Wang[3], Dinggang Shen[1]
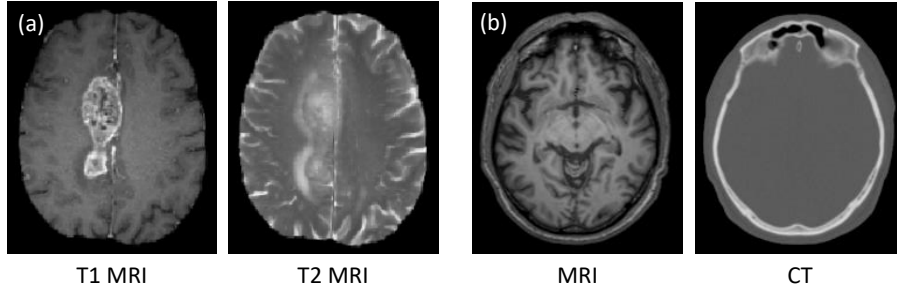
Fig. 1. Three pairs of corresponding source (left) and target (right) images from the same subjects. (a) shows a pair of T1 MRI/T2 MRI brain tumor images; (b) shows a pair of MRI/CT brain images.

is involved during acquisition). However, MRI is not directly related to tissue density information which is often required for radiotherapy planning or PET image reconstruction [9]. Based on the above observations, we argue that each modality is needed at different stages during disease diagnosis and treatment, while it is not easy to obtain them all in practice. To this end, it can be very beneficial to study the solutions that synthesizing modality of interest (or target) from the available source modalities. It provides the modality data required for many clinical trials without additional cost/risk of performing the actual acquisition.

Many researchers have tried to directly synthesize high-quality demanded medical modality images [6,15]. However, it is quite challenging due to the following possible problems. Firstly, the mapping from the source modality to the target modality (or its inverse) is typically complex and ill-posed [7]. Moreover, different modalities may show quite different image appearances, e.g., MRI and CT as shown in Fig. 1(b). Furthermore, certain regions in the image have totally different image contrast and even different shapes, for instance, the tumor regions in Fig. 1(a). Despite all these challenges, there are potential connections between the two modalities if we observe deep enough. That is, the mapping from source modality to target modality should be highly non-linear so that it can bridge the significant appearance gap between the two modalities.

Convolutional neural network (CNN) provides a natural way for learning highly non-linear mapping because of the multiple layer settings. In the particular case of image synthesis, Dong et al. [2] proposed to use CNNs for single image super-resolution. Li et al. [10] applied a similar deep learning model to estimate the missing PET image from the MRI data of the same subject. Huang et al. [8] proposed to simultaneously conduct super-resolution and cross-modality medical image synthesis by the weakly-supervised joint convolutional sparse coding. Nie et al. [15] proposed supervised adversarial learning framework with gradient difference loss to synthesize CT from MRI.

Although the training of the above mentioned image synthesis methods could achieve good performances in most cases, they cannot obtain reasonable results in certain situations, such as tumor regions when dealing with disease diagnosis shown in Fig. 1(a), since it does not consider samples or regions that are difficult though very important to process. The deep reason behind is that the training of the network tends to be dominated by samples/regions that are easy to synthesize, i.e., normal tissue regions. This easy-to-synthesize sample/region dominance phenomenon often occurs in medical image synthesis tasks due to the irregular tumor/lesion distribution. These tumor/lesion regions may be ignored owing to the relatively small size in the whole image, although they are the most important biomarkers for clinical diagnosis. Thus, it is quite important to develop a method that could better model the lesion/tumor regions in medical image synthesis tasks.

In this work, we propose a dual-discriminator adversarial learning framwork with difficult-region-aware mechanism to solve the above mentioned issues. Specifically, besides the regular CNN-based discriminator, we also propose a dense FCN as the local-discriminator to model the easy-or-hard extent for the synthesis task. More importantly, we further propose a difficult-region-aware mechanism to better model the hard-to-synthesis regions (i.e., tumor regions). Experimental results demonstrate that the proposed method can effectively synthesize target images with much better modeling capacity on the hard-to-synthesize regions. To the best of our knowledge, this is the first work that tries to address the hard-to-synthesis regions on cross-modality image synthesis tasks.

2   Methods

To address the above mentioned problems and challenges, we propose a deep convolutional adversarial network framework by adversarially training the generator (UNet) via dual discriminators: a CNN as the global discriminator and an FCN as the local discriminator, the framework of which is shown in Fig. 2. First, we propose a basic UNet generator to estimate the target image from the corresponding source image. Note that we adopt 2.5-D operations to better model the spatial mapping and thus could alleviate the discontinuity problem across 2D slices. Second, we utilize the adversarial learning strategy [4] for the designed image synthesis network, where two additional discriminator networks are modeled. The first discriminator (global one) urges the generator's output to be similar with the ground-truth target image perceptually. The second discriminator (local one) compares the local regions of the synthetic target images to those of the real target images, so that we can obtain local similarities between the synthetic and real images and thus further improve the hard-to-predict regions. At the testing stage, an input source image is first partitioned into overlapping patches, and, for each patch, the corresponding target is estimated by the generator. Then, all generated target patches are merged into a single image to complete the source-to-target synthesis by averaging over overlapping target
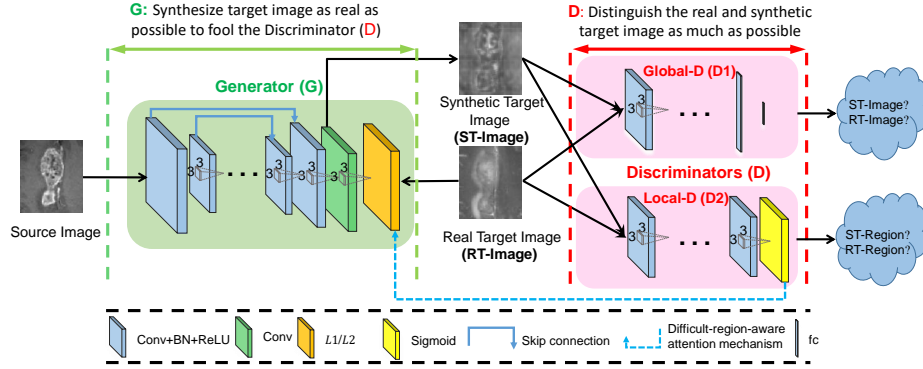
Fig. 2. Architecture used in the deep supervised generative adversarial setting to synthesize target image. This framework contains one generator and two discriminators. A difficult-region-aware attention mechanism is also included in the framework.

regions. In the following, we will describe in detail the proposed medical image synthesis framework.

2.1   Supervised Generative Adversarial Network

As mentioned above, we propose a supervised deep convolutional adversarial framework, which is inspired by the recent popular generative adversarial networks (GANs) [4], to complete the source-to-target synthesis as shown in Fig. 2. The components in Fig. 2 will be introduced in the following paragraphs.

2.2   UNet for Medical Image Synthesis

UNet, an evolutionary version of fully convolutional networks (FCN) which incorporates the high-resolution (low layers) feature maps and highly-semantic (high layers) feature maps to increase the localization accuracy, is widely used for segmentation and reconstruction in both computer vision and medical image analysis [18,3,6,14,16]. It can preserve spatial information in local neighborhood of the image space and is also much faster compared to CNN at the testing stage. In this paper, we adopt UNet to implement the image generator to perform the medical image synthesis task since UNet would potentially alleviate the loss of resolution compared to FCN.

As mentioned in the Introduction section, typically an L_1/L_2 loss is used to train the model as shown in Eq. 1.

$$L_G(X, Y) = \|Y - G(X)\|^p \tag{1}$$

where Y is the ground-truth target image, and $G(X)$ is the generated target image from the source image X by the Generator network G and p is 1 or 2.

2.3   Adversarial Learning for Medical Image Synthesis

To make the generated target images better perceptually [4,17,15], we propose to use adversarial learning to improve the performance of UNet. Instead of using classical CNN-based discriminator to globally evaluate how well the images are generated, we propose using an additional dense FCN-based local discriminator to capture the quality of the synthetic images for local regions.

Specifically, our networks include 1) the generator for estimating the target image and 2) the global discriminator for distinguishing the real target image from the generated one and a local discriminator distinguish synthetic and real images in voxel-level, as shown in Fig. 2. The generator network G is an FCN as described above. The global discriminator network $D1$ is a CNN, following the popular Wasserstein GAN [1,5]. The local discriminator is a FCN with a dense output, each element of which corresponds to the probability of how well the local region around this voxel is synthesized. It is north noting that global-D and local-D are independent. $D1$ and $D2$ are trained simultaneously, while G are iteratively updated with these two discriminators.

Global Adversarial Learning: The classical GAN [4] may lead to training failure due to the limitation of the JensenShannon divergence (JS-divergence) [1]. To address this issue, WGAN [1] uses the Wasserstein distance instead of the JS-divergence to compare data distributions. Later, gradient penalty strategy(WGAN-GP) [5] is further proposed to replace the gradient clipping in the WGAN.

In our study, we follow the WGAN-GP to form the global discriminator ($D1$). Concretely, the loss function for $D1$ can be defined as:

$$L_{D1}(X,Y) = E_x[D1(G(X))] - E_Y[D1(Y)] + \lambda E_{\hat{X}}[(\left\|\nabla_{\hat{X}}D1(\hat{X})\right\|_2 - 1)^2] \quad (2)$$

where X is the source input image, Y is the corresponding target image, $G(X)$ is the estimated image by the generator, and $\hat{X}$ is uniformly sampled along straight lines among synthetic and real samples; and λ is a constant weighting hyper-parameter.

On the other hand, the global adversarial loss term used to train G is defined as Eq. 3.

$$L_{G\_ADV1}(X,Y) = -E_x[D1(G(X))] \quad (3)$$

With the above equations, $D1$ can globally distinguish the real target data and the synthetic target data synthesized by G. At the same time, G aims to produce more realistic target images and to confuse $D1$. The architecture of $D1$ follows the suggestions in [5,1].

Local Adversarial Learning: The training objective of the confidence network is the summation of binary cross-entropy loss over the image domain, as shown in Eq. 4. Here, we use G and $D2$ to denote the generator and local-D networks, respectively.

$$L_{D2}(\mathbf{X}, \mathbf{Y}; \theta_{\mathbf{D2}}) = L_{BCE}(D2(\mathbf{Y}, \theta_{\mathbf{D2}}), \mathbf{1}) + L_{BCE}(D2(G(\mathbf{X}), \theta_{\mathbf{D2}}), \mathbf{0}), \quad (4)$$

where

$$L_{BCE}\left(\widehat{\mathbf{Q}}, \mathbf{Q}\right) = -\sum_{h=1}^{H}\sum_{w=1}^{W} Q_{h,w}\log\left(\widehat{Q}_{h,w}\right) + (1 - Q_{h,w})\log\left(1 - \widehat{Q}_{h,w}\right) \quad (5)$$

where $\mathbf{X}$ and $\mathbf{Y}$ represent the input data and its corresponding real target image, respectively. $\theta_{\mathbf{D2}}$ is network parameters for the local-D network.

For training the generator network, besides the L_1/L_2 loss defined in Eq. 1 and the global adversarial learning loss in Eq. 6, the local adversarial loss ("ADV") to improve G and fool $D2$ can be defined by Eq. 6:

$$L_{ADV2}\left(\mathbf{X}, \theta_{\mathbf{G}}\right) = L_{BCE}\left(D2\left(G\left(\mathbf{X}; \theta_{\mathbf{G}}\right)\right), \mathbf{1}\right) \quad (6)$$

The training of the two networks is performed in an alternating fashion. First, D is updated by taking a mini-batch of real target data and a mini-batch of generated target data (corresponding to the output of G). Then, G is updated by using another mini-batch of samples including sources and their corresponding ground-truth target images.

2.4   Region-attention based Adversarial Difficulty Learning

Due to the inhomogeneous characteristics and irregular distribution of the medical images, certain region of the images are usually more difficult to well synthesize. As a consequence, it is quite desired to build a model that can better model the hard-to-prediction regions. Since the local discriminator could provide the local confidence information about how well each region is synthesized, we can thus pay more attention on the hard-to-predict regions (e.g., lesion regions) so that these regions can be better modeled. To this end, we propose a adversarial difficulty-aware attention mechanism to better represent the easy-or-hard information. Specifically, we design a difficulty-aware L_1/L_2 loss using region-level attentions from the adversarial local confidence map.

The voxel-level difficulty-aware attention from the confidence map (M) is formulated (based on Eq. 1) in Eq. 7:

$$L_{AttG}(X, Y) = F \odot \|Y - G(X)\|^p \quad (7)$$

where $\odot$ is the element-wise multiplication and

$$F = (1 - M)^\beta \quad (8)$$

where β is the voxel-level attention parameter. Note, F here works as a scaling factor, which largely suppresses the contribution of easy-to-synthesize regions to the training loss and emphasize the hard-to-synthesize regions.

With the difficult-region-aware L_1/L_2 loss in Eq. 7, we can pay more attention in the less confidently (i.e., hard-to-predict) regions and thus better model them (e.g., tumor or lesion regions). And this adversarial difficulty-region-aware attention mechanism provides an opportunity to use voxel-wise focal loss in regression context.

Total Loss for Training Generator: To this end, the total loss for training generator includes the attention based L_1/L_2 loss, the global adversarial loss, and the local adversarial loss, which can be expressed as Eq. 9.

$$L_G = L_{AttG} + \lambda_1 L_{G\_ADV1} + \lambda_2 L_{G\_ADV2} \tag{9}$$

The above training loss could encourage G to generate target images following several constraints, i.e., besides the voxel-wise correspondence, it also needs to fool the discriminators both globally and locally.

2.5   Training Details

Shown in Fig. 2, this generator takes a source image as the input, and tries to generate the corresponding target image with a typical UNet. The discriminator $D1$ is a typical CNN including three stages of convolution, BN, ReLU and max pooling, followed by one convolutional layer and three fully connected layers where the first two use ReLU as activation functions. The filter size is 3×3, the numbers of the filters are 32, 64, 128, respecitvely, and 256 for the convolutional layers, and the numbers of the output nodes in the fully connected layers are 512, 128 and 1, respectively. The dense discriminator $D2$ is a typical FCN with three down-sampling and three up-sampling stages, followed by a convolutional layer with sigmoid activation. In both of the discriminators, we apply the spectral normalization [13] for all the layers except the last one.

We have randomly extracted source domain patches of size $5 \times 144 \times 144$, along with their corresponding target domain patches of size $1 \times 144 \times 144$, using the same center points, as the paired training samples. All networks were trained using the Adam optimizer, the initial learning rate for G is set as 5×10^{-3}, that of $D1$ is set as 1×10^{-4}, and that of $D2$ is set as 1×10^{-3}. Note that we decrease the learning rate with a rate of 0.5 for G, 0.2 for both $D1$ and $D2$ every 2 epochs. The mini-batch size is set to be 10. The generator was trained using $\lambda_1 = 0.05, \lambda_2 = 0.1$. The code is implemented using the pytorch library[1], and it will be publicly released upon acceptance.

3   Experiments and Results

We choose the BRATS dataset to evaluate our proposed method, which is a publicly available dataset of MRI from brain tumor patients [12]. A total of 354 pairs of T1 MRI and T2 MRI were assembled, where 200 subjects were used for training and 60 subjects were used for validation, and the rest 94 subjects were reserved for testing (Note, the dataset is randomly partitioned). The BRATS dataset is acquired under different scanning protocols on separate sites. Thus, the image sizes and resolutions are different across subjects.

To demonstrate the advantage of the proposed method in terms of prediction accuracy, we compare it with three widely-used approaches: atlas-based

---

[1] https://github.com/pytorch/pytorch

method [19], FCN [11], UNet [6], and sGAN [15] (For simplicity and fair comparison, we have removed auto-context refinement and use UNet as the generator). The experiments for the BRATS dataset are only once following the partition mentioned before. And the experiments for the brain MRI-to-CT dataset are done in a leave-one-out fashion. The evaluation metric is the mean absolute error (MAE) and the peak signal-to-noise ratio (PSNR).

3.1   Impact of Proposed Dual-Discriminator Strategy

To show the contribution of the proposed dual-discriminator strategy, we conduct comparison experiments between the sGAN with global discriminator (sGAN-1) and sGAN with local discriminator (sGAN-2) and the proposed dual-discriminator strategy (sGAN-dual) on the BRATS dataset. The PSNR values are 27.3dB, 27.2dB and 27.6dB for these three strategies, respectively. Note that these results are achieved with the ordinary L_1 loss for the generator. Actually, besides the explanation from local and global adversarial constraints, we can explain it from another view, i.e., the adversarial gradient vanishing issue can be greatly alleviated due to the dual-discriminator system which can thus better guarantee the adversarial learning.
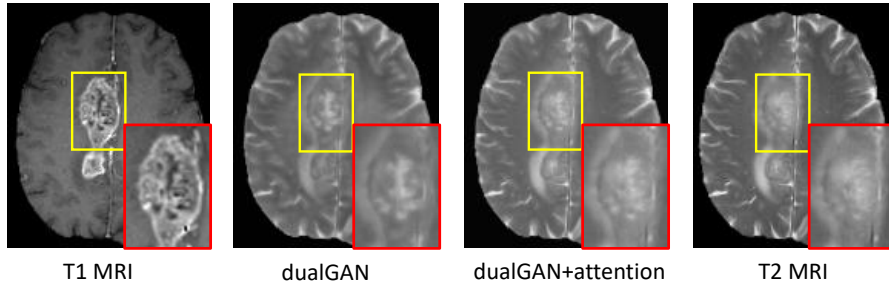
3.2   Impact of Difficult-Region-Aware Attention Mechanism

To show the impact of the proposed difficult-region-aware attention mechanism, we first conduct experiments to compare the performance without this mechanism and with this mechanism on the BRATS dataset, and the experimental results indicate that the performance could be improved by 0.2dB in terms of PSNR using the proposed attention mechanism. To further investigate the impact of the proposed mechanism, we focus on evaluating the synthesis performances only on tumor regions. By using the segmentation ground truth maps on this dataset, we have computed the PSNR on the testing sets only on the tumor regions, which indicates that the PSNR on tumor regions is improved by 0.6dB in average.

   We also visualize results in Fig 3, where the leftmost image is the input T1 MRI, and the rightmost image is the ground-truth T2 MRI. We can clearly see that the generated data using the proposed difficult-region-aware attention mechanism (i.e., 'dual-D+attention') could recover much more details compared to the one without this mechanism (i.e., 'dual-D'), especially for the tumor regions, which further demonstrate the effectiveness of our proposed difficult-region-aware attention mechanism.

   To better understand why the difficult-region-aware mechanism works, we also analyze the confidence map generated by the local discriminator (i.e., $D2$). We find that the tumor regions are evaluated to be poorly synthesized, thus, more attention will be propagated to the tumor regions in the generator network, as a consequence, these regions are then better modeled during training.

T1 MRI   dualGAN   dualGAN+attention   T2 MRI

Fig. 3. Visual comparison for impact of the proposed difficult-region-aware attention mechanism.

3.3   Comparing with Other Methods

To qualitatively compare the synthetic target image by different methods, we visualize the generated target image with the ground-truth target image in Fig. 4. We can see that the proposed algorithm can better preserve the continuity, coalition and smoothness in the synthetic results, since it uses both global and local adversarial learning constraints in the image patch as discussed in Section 2.1. More importantly, the tumor region of generated T1 MRI can recover much more details than other methods, and thus looks much closer to the real T2 MRI compared to all other methods. We argue that this is due to the difficult-region-aware attention mechanism which reweight more on the recognized hard-to-synthesis regions, i.e., tumor regions.

We also quantitatively compare the predicted results in Table 1 in terms of PSNR and MAE. Our proposed method outperforms other methods in both metrics, which is consistent with the visual comparison and further demonstrates the advantage of our framework.

Table 1. Average MAE and PSNR on 94 **Table 2.** Average MAE and PSNR on testing subjects from the BRATS dataset. 16 subjects from the brain dataset.

Method	MAE	PSNR
FCN	121.3(18.6)	22.7(2.3)
UNet	107.6(**14.1**)	26.0(**1.4**)
sGAN	100.3(14.7)	26.6(1.5)
Ours	**92.4**(14.2)	**27.3(1.4)**

Method	MAE	PSNR
FCN	24.4(15.1)	22.7(3.2)
UNet	21.8(12.8)	26.7(2.1)
sGAN	20.4(11.2)	27.3(**1.7**)
Ours	**18.6(10.3)**	**28.1**(1.8)

3.4   Validation on Another Brain Dataset

To show the generalization ability of the proposed method, we also validate the proposed method on the brain dataset, which was acquired from 16 sub-
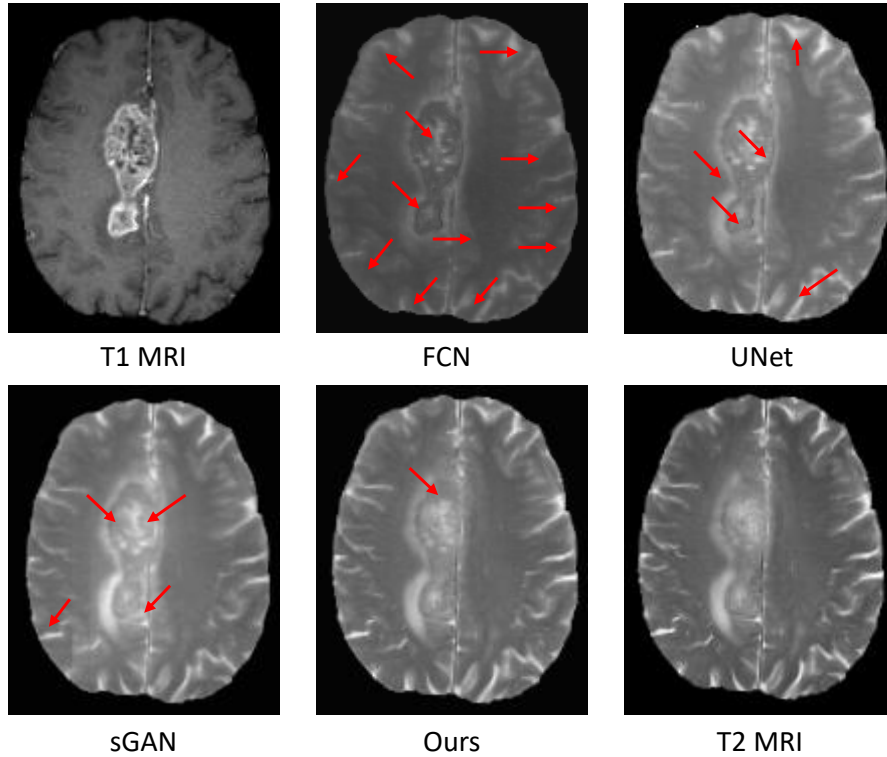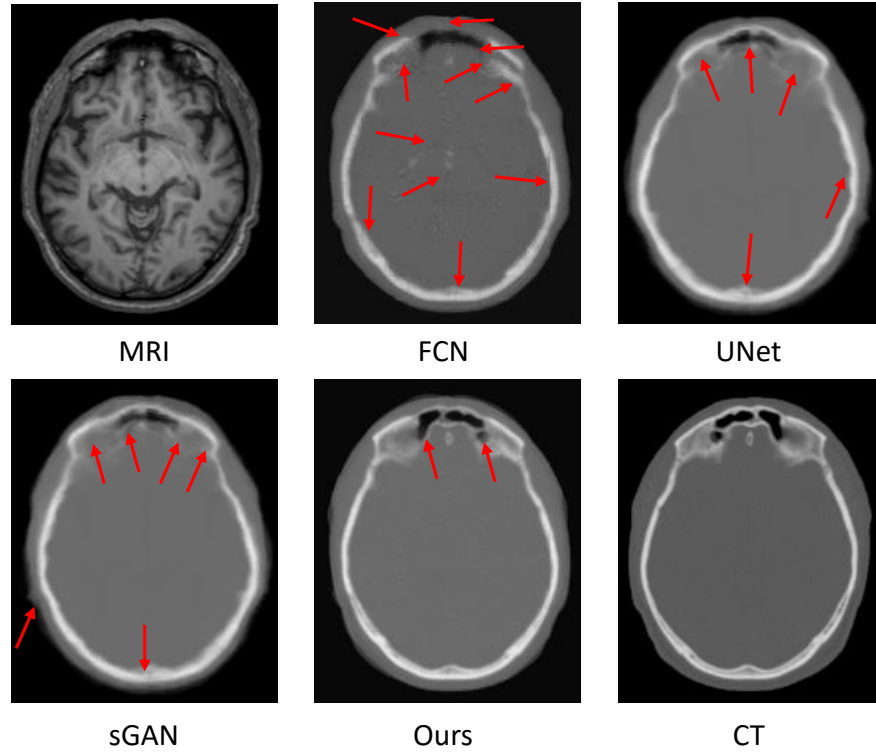
Fig. 4. Visual comparison of MR image, the estimated CT images by our method and other competing methods, and the ground-truth CT image for the typical brain tumor cases. Red arrows mean poorly synthesized regions.

jects with both MRI and CT scans in the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (see `www.adni-info.org` for details). A typical example of preprocessed CT and MR images is given in Fig. 1. The qualitative comparison results with different methods are shown in Fig. 5, and the quantitative comparison results are shown in Table. 2. Obviously, our proposed method can work better than the state-of-the-art methods, which demonstrate that our proposed method can be generalized to more datasets. More importantly, the proposed method could model much better details after clearly observing the visual results.

4   Conclusions

We developed a dual discriminator based adversarial learning framework, including a global one modeling the overall evaluation and a local one modeling the region-wise evaluation, to solve several medical image synthesis tasks. Moreover, we proposed a difficult-region-aware attention mechanism to better

Fig. 5. Visual comparison of MR image, the estimated CT images by our method and other completing methods, and the ground-truth CT image for the typical brain case. Red arrows mean poorly synthesized regions.

model the hard-to-predict regions (e.g., tumor regions). We have applied our proposed model on two tasks, i.e., to predict T2 MRI from their corresponding T1 MRI and to predict brain CT images from their corresponding MR images. The experimental results demonstrate that our proposed method could outperform three state-of-the-art methods. Furthermore, the proposed difficult-region-aware attention mechanism could indeed help improve the hard-to-synthesis regions. Lastly, the proposed method could also be applied to other related medical image synthesis tasks.

References

1. Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein gan. *arXiv preprint arXiv:1701.07875*, 2017.
2. Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *ECCV*, pages 184–199. Springer, 2014.

3. Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE TPAMI*, 38(2):295–307, 2016.
4. Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *NIPS*, pages 2672–2680, 2014.
5. Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C Courville. Improved training of wasserstein gans. In *NIPS*, pages 5767–5777, 2017.
6. Xiao Han. Mr-based synthetic ct generation using a deep convolutional neural network method. *Medical Physics*, 44(4):1408–1419, 2017.
7. Alaa A. Hefnawy. *Super Resolution Challenges and Rewards*, pages 163–206. Atlantis Press, Paris, 2010.
8. Yawen Huang, Ling Shao, and Alejandro F Frangi. Simultaneous super-resolution and cross-modality synthesis of 3d medical images using weakly-supervised joint convolutional sparse coding. *arXiv preprint arXiv:1705.02596*, 2017.
9. Paul E Kinahan, DW Townsend, T Beyer, and D Sashin. Attenuation correction for a combined 3d pet/ct scanner. *Medical physics*, 25(10):2046–2053, 1998.
10. Rongjian Li, Wenlu Zhang, Heung-Il Suk, Li Wang, Jiang Li, Dinggang Shen, and Shuiwang Ji. Deep learning based imaging data completion for improved brain disease diagnosis. In *MICCAI*, pages 305–312. Springer, 2014.
11. Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *CVPR*, pages 3431–3440, 2015.
12. Bjoern H Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE TMI*, 34(10):1993, 2015.
13. Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida. Spectral normalization for generative adversarial networks. *arXiv preprint arXiv:1802.05957*, 2018.
14. Dong Nie, Xiaohuan Cao, Yaozong Gao, Li Wang, and Dinggang Shen. Estimating ct image from mri data using 3d fully convolutional networks. In *DLMIA*, pages 170–178. Springer, 2016.
15. Dong Nie, Roger Trullo, Jun Lian, Caroline Petitjean, Su Ruan, Qian Wang, and Dinggang Shen. Medical image synthesis with context-aware generative adversarial networks. In *MICCAI*, 2017.
16. Dong Nie, Li Wang, Yaozong Gao, Jun Lian, and Dinggang Shen. Strainet: Spatially varying stochastic residual adversarial networks for mri pelvic organ segmentation. *IEEE transactions on neural networks and learning systems*, 30(5):1552–1564, 2018.
17. Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.
18. Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, pages 234–241. Springer, 2015.
19. Tom Vercauteren, Xavier Pennec, Aymeric Perchant, and Nicholas Ayache. Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage*, 45(1):S61–S72, 2009.