https://doi.org/10.26599/AIR.2024.9150035 CAAI Artificial Intelligence Research

ADGAN: Adaptive Domain Medical Image Synthesis Based on Generative Adversarial Networks

Liming Xu[1, 2] ✉, Yanrong Lei[1], Bochuan Zheng[1], Jiancheng Lv[2], and Weisheng Li[3]

ABSTRACT

Multimodal medical imaging of human pathological tissues provides comprehensive information to assist in clinical diagnosis. However, due to the high cost of imaging, physiological incompatibility, and the harmfulness of radioactive tracers, multimodal medical image data remains scarce. Currently, cross-modal medical synthesis methods can generate desired modal images from existing modal images. However, most existing methods are limited to specific domains. This paper proposes an Adaptive Domain Medical Image Synthesis Method based on Generative Adversarial Networks (ADGAN) to address this issue. ADGAN achieves multidirectional medical image synthesis and ensures pathological consistency by constructing a single generator to learn the latent shared representation of multiple domains. The generator employs dense connections in shallow layers to preserve edge details and incorporates auxiliary information in deep layers to retain pathological features. Additionally, spectral normalization is introduced into the discriminator to control discriminative performance and indirectly enhance the image synthesis ability of the generator. Theoretically, it can be proved that the proposed method can be trained quickly, and spectral normalization contributes to adaptive and multidirectional synthesis. In practice, comparing with recent state-of-the-art methods, ADGAN achieves average increments of 4.7% SSIM, 6.7% MSIM, 7.3% PSNR, and 9.2% VIF.

KEYWORDS

generative adversarial networks; multimodal medical images; adaptive domain; latent shared representation

Multimodal medical images play a crucial role in clinical treatment, and the different modalities provide complementary information to enhance our understanding of various physiological processes and pathological states. The comprehensive utilization of these different information will help medical workers to understand the disease more comprehensively and provide more accurate diagnosis and treatment options for patients. Magnetic resonance imaging (MRI) highlights different tissue features by adjusting the pulse sequence, while Computed Tomography (CT) uses a precise X-ray beam to quickly slice a region of the body. Positron Emission Computed Tomography (PET) marks the lesion area with radionuclides to reflect its metabolic process[1]. However, it is difficult to obtain complete contrast information from MRI due to physiological incompatibility during long-term acquisition. Additionally, PET imaging requires the injection of a radioactive tracer into patients, which causes irreversible damage and additional economic burden.

Multimodal medical image synthesis, which attempts to synthesize target modal images from acquired source images and make synthesized images close to real images in terms of perception and quality. These methods can be generally divided into handcrafted and deep learning-based methods; in this paper, we mainly focus on the latter. Deep learning-based methods use deep neural networks to learn nonlinear mappings between different modalities to achieve high-quality image synthesis and produce promising results. At this stage, the BPGAN[2] and Bi-MGAN[3] design bidirectional mapping between CT-MRI and T1-T2 MR images while preserving pathological structure and information.

Despite the progress made by the existing deep learning-based methods in pathological structure and physician distinction, there are still two main challenges. (1) Almost all the existing approaches can only achieve mapping between two fixed domains, and (2) they have low robustness, which means that one mapping requires one predictor and that the trained predictor within a specific domain cannot be transferred to other domains. As shown in Figs. 1a–1c, well-trained MR predictors made by the BPGAN, Bi-MGAN, and our Adaptive Domain Medical Image Synthesis Method based on Generative Adversarial Networks (ADGAN) can perform satisfactory cross-modal synthesis. However, the BPGAN, and Bi-MGAN yield unsatisfactory results when generating CT from MR images by using the same well-trained predictor, as shown in Figs. 1d and 1e, which means that a well-trained MR predictor cannot be directly transferred to achieve MRI→CT synthesis. In contrast, the ADGAN achieves superior results with a single generator compared to BPGAN and Bi-MGAN, which use separate generators for each modality.

To address limitations in mapping between fixed domains and enhance robustness while considering the consistency of pathological information between source and real target images, we propose ADGAN to achieve adaptive and multidirectional cross-modal medical image synthesis without training multiple generators. The main contributions of this study can be summarized as follows.

- We propose an adaptive domain synthesis method based on

1 School of Computer Science, China West Normal University, Nanchong 637002, China
2 College of Computer Science and Technology, Sichuan University, Chengdu 130012, China
3 College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
Address correspondence to Liming Xu, xulimmail@gmail.com

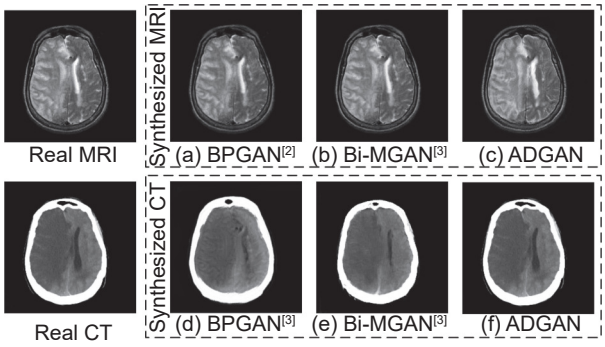

Fig. 1 Bidirectional and adaptive synthesis.

generative adversarial networks to achieve multidirectional medical image generation without using multiple generators. High-quality images under multiple target domains are synthesized by one generator that learns latent shared features across multiple domains, which represents the first attempt in the medical image cross-modal synthesis community.

- We combine Densely Connected Blocks (DenseBlock) and Residual Blocks (ResBlock) to construct a generator to reuse shallow features and preserve low-level visual features and high-level semantic features in the source domain. Then, we demonstrate that the generator has obvious advantages in terms of training speed.
- We introduce spectral normalization into the adaptive multidirectional discriminator to control the Lipschitz upper bound, which balances the performance between the generator and the discriminator and enables the generator to stably fit the real medical image distribution. Additionally, a detailed analysis is provided to show the superiority of adaptive domain synthesis.
- The qualitative and quantitative results show that the proposed approach outperforms recent state-of-the-art cross-modal synthesis methods. Compared with recent state-of-the-art methods, our method achieves average improvements of 4.7% in Structural Similarity (SSIM), 6.7% in Multiscale Structural Similarity (MSIM), 7.3% in Peak Signal-to-Noise Ratio (PSNR), and 9.2% in Visual Quality Fidelity (VIF).

The remainder of this article is organized as follows. Section 1 introduces related work on medical image cross-modal synthesis. Section 2 proposes our ADGAN for medical image adaptive domain synthesis. Section 3 provides algorithm optimization and theoretical analysis. Section 4 presents the extensive results, with concluding remarks and suggestions for future work summarized in the last section.

1 Related Work

The existing cross-modal medical image synthesis methods generally include unidirectional and bidirectional synthesis methods as follows.

1.1 Unidirectional synthesis method

Unidirectional synthesis always synthesizes images within a specific domain from another domain, e.g., CT can be synthesized from the MRI domain. Considering the progress made by Deep Neural Networks[4] (DNNs) and image registration, Ref. [5] develops a DNN-based approach to produce accurate CT images from conventional and single-sequence MR images in near real time. Similarly, multiplanar DCNNs[6] have been proposed, and the feasibility of MRI-based proton therapy planning for the brain has been verified by assessing the range shift error within the clinical acceptance threshold. Considering the advantages of Fully Convolutional Networks (FCNs)[7], several researchers have proposed estimating PET from CT scans[8] and synthesizing multicontrast MR images[9]. In addition, Ref. [10] combines FCN and U-Net to synthesize CT from MR in various cases with extensive results.

Generative Adversarial Networks (GANs)[11] have been applied in many fields, including natural and medical image processing[12], due to their powerful fitting ability. GANs take a low-dimensional sample as input and produce high-resolution output with fine details[13]. GAN-based methods demonstrate promising outcomes in enhancing the resolution and quality of medical images, which are critical for accurate diagnosis and clinical decision-making. Based on this, sequential GAN[14], which is an unconditional GAN model, generates bimodal images with pairwise relationships from randomly sampled noise. Then, SkrGAN[15] introduces sketch prior constraints to preserve fine structures in medical images. Similarly, Ref. [16] further improves the synthesis performance by adding auxiliary conditional constraints. mustGAN[17] fuses multiple one-to-one streams and one joint multiway stream to achieve multicontrast MR image synthesis. Reference [18] utilizes a conditional adversarial spatiotemporal encoder that extracts spatial and temporal features to predict tumor growth.

Considering edge information, EaGAN[19] uses the Sobel edge detection operator to integrate edge information and construct an edge-aware GAN for cross-modal MR image synthesis. Similarly, EPIMFGAN[20] improves edge detail using an iterative multiscale fusion module. Then, MGMGAN[21] introduces a gate mergence mechanism to automatically learn the weights of different modalities at different locations to optimize multiscale fusion. ARGAN[22] synthesizes high-quality standard-dose Single Photon Emission Tomography (SPET) from low-dose PET and uses spectral regularization to constrain the consistency between synthesized and real images in the frequency domain. Based on the pix2pix model, Ref. [23] builds a dermoscopic image synthesis algorithm and fused deep and shallow features to enhance synthetic texture details. DiCyc[24] achieves precise alignment for cross-modal synthesis by filtering out region-specific deformations upon cyclic consistency. Due to spatial inconsistency, 2D GANs[25] are not always synthesized satisfactorily. Thus, some approaches[26–28] combined 3D GANs to model 3D spatial information and solve the problem of cross-slice discontinuities.

1.2 Bidirectional synthesis method

Although unidirectional methods can be used to synthesize desired modal images, they suffer from low flexibility in inverse mapping. Compared with unidirectional synthesis, the bidirectional synthesis method can obtain abundant multimodal images in a more flexible way. Thus, several bidirectional medical image synthesis methods have also been proposed. cGAN[29] preserves source image details using cyclic consistency loss for T1-w to T2-w bidirectional synthesis. BPGAN[2] achieved bidirectional prediction between CT-MR images by using a multigenerative multiadversarial mechanism with the constraint of pathological information and a reduction in generalization errors. Similarly, Bi-MGAN[3] utilizes a multigenerative multiadversarial mechanism and fuse depth and hand-crafted features to maintain anatomical structure to achieve bidirectional prediction between T1-w and T2-w. BMGAN[30] achieves bidirectional synthesis of brain MR-PET images by simulating bidirectional mapping of the high-dimensional latent space of MR images to PET images instead of bidirectional mapping between image spaces.

Unidirectional synthesis methods perform well in specific

domains with reasonable training strategies or auxiliary conditions. Nevertheless, these methods struggle to handle complex tasks with multiple mapping directions. Bidirectional synthesis methods, which are mainly based on GANs, overcome the limitation of mapping direction by using cycle consistency to achieve bidirectional mapping between two domains. However, both unidirectional and bidirectional synthesis methods are limited to mapping between two domains, and they cannot overcome the limitation that one generator learns only one specific domain. That is, if there are N domains, then $N(N-1)$ generators must be trained, which greatly limits the scalability.

Inspired by facial attribute editing, which learns latent shared features, we propose ADGAN, which utilizes one generator to synthesize multiple domain images and fit the manifold across multiple modalities to achieve adaptive multidirectional medical image synthesis. The generator incorporates dense connected and residual blocks and reuses shallow and deep features to preserve semantic and pathological information. Then, the hybrid generator can be trained faster than separated dense or residual blocks. In addition, we introduce spectral normalization into an adaptive, multidirectional discriminator to control the Lipschitz upper bound, which balances the performance between the generator and the discriminator and enables the generator to fit the real medical image distribution stably.

The most closely related methods to our own are StarGAN[31] and StarGAN V2[32], which were originally designed for style translation. These two models can also achieve multidomain image synthesis by learning deterministic mappings for attribute labels. However, simple style translation cannot guarantee pathological consistency, and missing or distorted pathological information can even mislead clinical diagnosis. In this regard, we consider the label input with domain code input to achieve multiple corresponding domain generation. Then, we propose domain feature reconstruction loss to force the generator to synthesize targeted images with both style and pathology constraints. Perceptual loss is designed to ensure structural consistency in shallow manifold space since the texture structure is crucial in clinical diagnosis.

2 Network Architecture and Loss Definition

In this section, we present the deep architecture and loss definition of the proposed ADGAN.

2.1 Network architecture

Different from the abovementioned StarGAN[31] and StarGAN V2[32], the ADGAN does not simply transfer the style of the target image to the source image but preserves the anatomical structure and pathological details in the target image. As shown in Fig. 2, ADGAN contains four networks, i.e., the mapping network F, feature encoder E, generator G and discriminator D. The latent code z is encoded into domain feature code $\tilde{s}_y$ through the mapping network F, and the reference image y is input into the feature encoder E to extract the style code $\tilde{s}$. During training, domain feature code $\tilde{s}_y$ and style code $\tilde{s}$ are alternately used as auxiliary inputs to the generator, we now provide a detailed description of each module in Section 2, and define the loss function using domain feature code as an auxiliary input.

2.1.1 Mapping network

The mapping network F maps randomly sampled latent code z into domain code $\tilde{s}_y$ with a target domain style. We use a multilayer perceptron to construct a mapping network with a total of eight fully connected layers per domain and use the ReLU activation function after the first to the seventh layers. Assuming that the source domain is X and the target domain is Y, the mapping network F is used to map the latent code z with random sampling into target domain feature code $\tilde{s}_y$. Then, $\tilde{s}_y$ is used as an auxiliary input to the generator to help the model learn the latent features of the target domain.

2.1.2 Feature encoder

The feature encoder E, which consists of six residual blocks and one fully connected layer, is used to extract the style code of the target domain. The residual blocks extract the high-dimensional features of the input image, and the fully connected layer linearly scales them into style code $\tilde{s}$. Note that the feature encoder reverses the encoding of the reference image y or the fake image into the latent space during training, and $\tilde{s}$ is used as an auxiliary input to the generator to guide the generation of the target image during validation.

2.1.3 Generator

The generator G, which contains an encoder and decoder

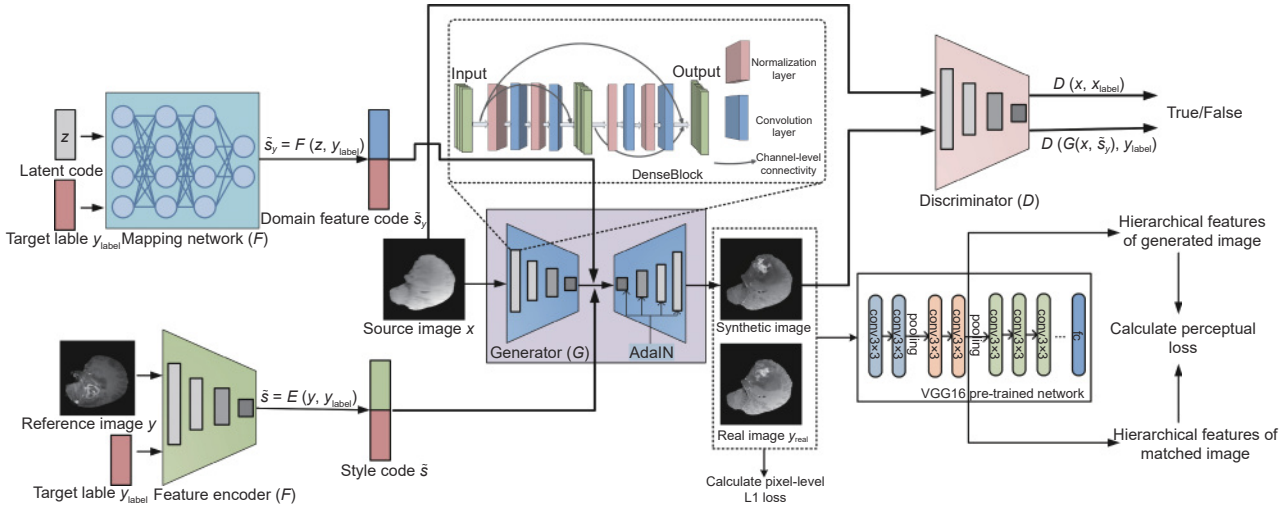

Fig. 2 Architecture and data flow of ADGAN, which contains a mapping network F, a feature encoder E, a generator G and a discriminator D. During training, domain feature code $\tilde{s}_y$ and style code $\tilde{s}$ iteratively alternate as auxiliary inputs to the generator, synthesized image and source image are input to the discriminator.

module, generates a synthesized image in the target domain with auxiliary input and the source image. The encoder mainly encodes the source image as a high-dimensional latent feature and contains three densely connected blocks and two residual blocks. The decoder reconstructs the high-dimensional latent features into a synthesized image that is desired to reside on the manifold of the target domain. During decoding, we use Adaptive Instance Normalization (AdaIN) to fuse content features in the source and target domains and map the fused features to the target image domain.

2.1.4 Discriminator

The discriminator D receives generated multiple synthesized images and target image labels and discriminates whether the synthesized medical images belong to the target domain. The discriminator consists of six residual blocks, and the instance normalization in the residual blocks is replaced by spectral normalization. It is shown later that introducing spectral normalization into an adaptive discriminator can also constrain the discriminative ability of D to balance the performance between the generator and discriminator, and further improve the performance of adaptive domain medical image synthesis.

2.2 Loss function

The goal of ADGAN is to generate pathological and perceptual target images close to the real image in multiple target domains. To achieve this goal, ADGAN uses adversarial loss, domain feature reconstruction loss, pixel-level L1 loss, and perceptual loss. Now we provide a detailed description of each loss. Explanations of all symbols are given in Table 1.

2.2.1 Adversarial loss

Define source domain X and source image x. The domain style code $\tilde{s}_y$ and generator G are required to generate a synthesized image $G(x,\tilde{s}_y)$ close to a real image y_{real}, while the discriminator D is to distinguish between this synthesized image $G(x,\tilde{s}_y)$ and a source image x. The adversarial loss of the discriminator D_{adv} and the generator G_{adv} is defined as follows:

$$D_{\text{adv}} = \boldsymbol{E}_x[\log(x)] + \boldsymbol{E}_{x,z}[\log(1 - D(G(x,\tilde{s}_y)))] \tag{1}$$

$$G_{\text{adv}} = \boldsymbol{E}_{x,z}[\log(1 - D(G(x,\tilde{s}_y)))] \tag{2}$$

where $\boldsymbol{E}[\cdot]$ denotes expectation, $\tilde{s}_y = F(z, y_{\text{label}})$ denotes domain feature code of the target domain and z is latent code for random sampling. y_{label} is domain label of target domain.

2.2.2 Domain feature reconstruction loss

To enable generator to learn latent shared features across multiple target domains, the domain feature reconstruction loss is embedded in conditional input of the auxiliary generation, i.e., domain feature code $\tilde{s}_y$, to further improve synthesis quality. The domain feature reconstruction loss is defined as

$$\mathcal{L}_{\text{sty}} = \boldsymbol{E}_{x,z}[||\tilde{s}_y - E(G(x,\tilde{s}_y), y_{\text{label}})||_1] \tag{3}$$

where E represents the feature encoder, which encodes the synthesized image $G(x,\tilde{s}_y)$ back into the latent space of the target domain, y_{label} represents the target domain label.

2.2.3 Pixel-level L1 loss

To ensure that the synthesized images $G(x,\tilde{s}_y)$ preserve the details and structure of the original images, L1 loss is incorporated to maintain spatial consistency of anatomical structures, It has been reported that L1 loss is more sensitive to outliers since it does not square the differences, which means that it places greater emphasis on edge and fine detail information[33]. The pixel-level L1 loss is used to make the synthesized image close to real images y_{real} in the target domain at the spatial level. Pixel-level L1 loss which reduce pixel-to-pixel intensity difference can be represented as

$$\mathcal{L}_{\text{L1}} = \boldsymbol{E}_{x,y_{\text{real}},z}[||y_{\text{real}} - G(x,\tilde{s}_y)||_1] \tag{4}$$

where y_{real} is paired image in target domain Y. $G(x,\tilde{s}_y)$ represents the corresponding synthesized image.

2.2.4 Perceptual loss

To maintain pathological invariance, i.e., synthesized image $G(x,\tilde{s}_y)$ should have the same pathological information as the real image y_{real}, the perceptual loss is constructed by using a pre-trained VGG16 model to extract feature maps in the depth space. The perceptual loss can be denoted as

$$\mathcal{L}_{\text{per}} = \boldsymbol{E}_{x,y_{\text{real}},z}[||V(y_{\text{real}}) - V(G(x,\tilde{s}_y))||_1] \tag{5}$$

where $V(\cdot)$ represents the pre-trained VGG16 network, where the hierarchical features are extracted after the fourth ReLU function activation operation.

2.2.5 Total loss function

We merge the above four loss functions together with hyper-parameter λ_{per}, λ_{sty}, and λ_{L1} to construct the whole optimization, which can be denoted as

$$\min_{G,F,E}\max_{D} \mathcal{L}_{\text{adv}} + \lambda_{\text{sty}}\mathcal{L}_{\text{sty}} + \lambda_{\text{L1}}\mathcal{L}_{\text{L1}} + \lambda_{\text{per}}\mathcal{L}_{\text{per}} \tag{6}$$

In this paper, $\lambda_{\text{per}} = 100$, λ_{sty} and λ_{L1} are set to be 1.

Table 1 Symbol definition.

Symbol	Definition	Symbol	Definition
G	Generator	x	Source image
D	Discriminator	y_{real}	Real image
E	Feature encoder	y_{label}	Target label
F	Mapping network	z	Latent code
V	VGG16 model	$\tilde{s}_y$	Domain feature code
X	Source domain	$G(x,\tilde{s}_y)$	Synthesized image
Y	Target domain	$\tilde{s}$	Style code

3 Algorithm and Analysis

ADGAN learns latent shared features to control synthesis across multiple domains, and further achieves adaptive and multidirectional medical images without using multiple generators. Following the previous works[2, 3, 32], and the Adam optimizer is used to alternately optimize and parameters. The learning algorithm of ADGAN is shown in Algorithm 1.

3.1 Adaptive synthesis analysis

In order to achieve adaptive synthesis across multiple target domains while preserving the pathological features in the target image, we design an adaptive mechanism that can be expressed as

Algorithm 1 ADGAN learning algorithm

Require: Source domain images X_k^a, target domain images Y_k^b, source domain label l^a, target domain label l^b, mini-batch sampling SA, batch size K, iteration count T, current iteration count t, learning rate lr, VGG16 network $V(\cdot)$.

Ensure: Generator G with parameters θ_g, mapping network F with parameters θ_f, feature encoder E with parameters θ_e.

Initialize: Randomly sample latent code z and discriminator parameters θ_d.

while $t < T$ **do**

$(x_k^a, y_k^b) \leftarrow SA(X_k^a, Y_k^b)$ {Sample mini-batch paried images.}

$\tilde{s}_b \leftarrow F(l^b, z)$ {Obtain domain feature code.}

$x_k^b \leftarrow G(x_k^a, \tilde{s}_b)$ {Synthesize fake images.}

$\theta_d \leftarrow \sigma_s(\theta_d)$ {Normalize discriminator using spectral normalization.}

{Calculate the gradient with respect to the parameter θ_d by using}:

$$\frac{\partial \mathscr{L}_D}{\partial \theta_d} = \frac{1}{k}\nabla\theta_d \sum_{k=1}^{K}\left[\log D\left(x_k^a, l^a\right) + \log\left(1 - D\left(x_k^b, l^b\right)\right)\right]$$

$\mathscr{L}_D \leftarrow \mathscr{L}_D + \mathrm{lr}\cdot\partial\mathscr{L}_D/\partial\theta_d$ {Update discriminator parameters θ_d.}

Calculate the gradient with respect to the parameter θ_g by using Eq. (11).

$\mathscr{L}_G \leftarrow \mathscr{L}_G - \mathrm{lr}\cdot\partial\mathscr{L}_G/\partial\theta_g$ {Update generator parameters θ_g.}

Calculate the gradient with respect to the parameter θ_f by using Eq. (12).

$\mathscr{L}_{sty} \leftarrow \mathscr{L}_{sty} + \mathrm{lr}\cdot\partial\mathscr{L}_{sty}/\partial\theta_f$ {Update mapping network parameters θ_f.}

{Calculate the gradient with respect to the parameter θ_e by using}:

$$\frac{\partial \mathscr{L}_{sty}}{\partial \theta_e} = \frac{1}{k}\nabla\theta_e \sum_{k=1}^{K}\log[-E_{\theta_e}(l^b, G(x_k^a, F(l^b, z))]$$

$\mathscr{L}_{sty} \leftarrow \mathscr{L}_{sty} + \mathrm{lr}\cdot\partial\mathscr{L}_{sty}/\partial\theta_e$ {Update feature encoder parameters θ_e.}

end while

$$\mathrm{AdaIN}(x_i, \tilde{s}_y) = \sigma(\tilde{s}_y)\left(\frac{x_i - \mu(x_i)}{\sigma(x_i)}\right) + \mu(\tilde{s}_y) \tag{7}$$

where x_i is the i-th layer feature extracted from the source image. The affine parameters can be obtained by calculating the mean $\mu(\cdot)$ and variance $\sigma(\cdot)$ of the domain feature code. Then, the source image is aligned to the target image by scaling and translating the normalized source image features. Essentially, AdaIN is a fusion of the i-th layer feature x_i of the source image and the feature $\tilde{s}_y$ of the target domain on multiple scales. Then, the decoder in the generator remaps the fused feature t to the image space and reconstructs the target image, which is recorded as:

$$G(x, \tilde{s}_y) = \mathrm{De}(\mathrm{AdaIN}(x_i, \tilde{s}_y)) \tag{8}$$

where $\mathrm{De}(\cdot)$ represents the decoder. From Eq. (8), it can be seen that whether the generator synthesizes an image with correct mapping relationship essentially replies on training decoder for decoding. Thus, the proposed ADGAN adopts the domain feature reconstruction loss to learn domain feature representation, and uses perceptual loss to decoding results close to the real image in the target domain. That is,

$$\mathscr{L}_1 = ||t - E(\mathrm{De}(\mathrm{AdaIN}(x_i, \tilde{s}_y)))||_1 \tag{9}$$

and

$$\mathscr{L}_2 = ||V(y_{\mathrm{real}}) - V(\mathrm{De}(\mathrm{AdaIN}(x_i, \tilde{s}_y)))||_1 \tag{10}$$

where E represents feature encoder and $V(\cdot)$ represents the pre-trained VGG16 network.

From Eqs. (9) and (10), we can conclude that a simple encoder-decoder structure can achieve adaptive domain synthesis when quantifying the clinical style of medical image by using mean and variance of the target images. i.e., the synthesis performance is not directly related to quantity of generators, we can construct one generator with encoder-decoder structure to learn latent shared representations across multiple domains, which provides a reasonable explanation for performance guarantee when replacing multiple generators with one generator in adaptive multi-directional medical image synthesis across multiple domains.

$$\begin{aligned}\frac{\partial \mathscr{L}_G}{\partial \theta_g} = \frac{1}{k}\nabla\theta_g \sum_{k=1}^{K}&\log\left[1 - D\left(G_{\theta_g}\left(x_k^a, F\left(l^b, z\right)\right), l^b\right)\right] + \\ &\log\left[F\left(l^b, z\right) - E\left(l^b, G_{\theta_g}\left(x_k^a, F\left(l^b, z\right)\right)\right)\right] + \\ &\log\left[y_k^b - G_{\theta_g}\left(x_k^a, F\left(l^b, z\right)\right)\right] + \\ &\log\left[V\left(y_k^b\right) - V\left(G_{\theta_g}\left(x_k^a, F\left(l^b, z\right)\right)\right)\right]\end{aligned} \tag{11}$$

$$\begin{aligned}\frac{\partial \mathscr{L}_{sty}}{\partial \theta_f} = \frac{1}{k}\nabla\theta_f \sum_{k=1}^{K}&\log\left[1 - D\left(G\left(x_k^a, F_{\theta_f}\left(l^b, z\right)\right), l^b\right)\right] + \\ &\log\left[F_{\theta_f}\left(l^b, z\right) - E\left(l^b, G\left(x_k^a, F_{\theta_f}\left(l^b, z\right)\right)\right)\right] + \\ &\log\left[y_k^b - G\left(x_k^a, F_{\theta_f}\left(l^b, z\right)\right)\right] + \\ &\log\left[V\left(y_k^b\right) - V\left(G\left(x_k^a, F_{\theta_f}\left(l^b, z\right)\right)\right)\right]\end{aligned} \tag{12}$$

3.2 Spectral normalization

It has been reported that the spectral normalization can control the Lipschitz constraint of discriminator, which balances the performance between generator and discriminator and indirectly improves generator[2]. Now, we show that it can also work in adaptive domain synthesis which generator and discriminator learn and model latent shared representations instead of learning specific-domain separately. Assume that the discriminator satisfies the γ-Lipschitz constraint, then for any latent manifold distribution p and q, we have

$$\frac{||L(p) - L(q)||_2}{||p - q||_2} \leqslant \gamma \tag{13}$$

In Eq. (13), Lipschitz constraint is satisfied when $\gamma = 1$. As mentioned before, the discriminator consists of six residual blocks, which can be denoted as

$$L^*(x) = \sum_{i=1}^{6} F(x_i, w_i) + x_0 \tag{14}$$

where $L^*(x)$ denotes the output before activation. x_0 is the initial input, and x_i and w_i are the input and corresponding weight matrix of the i-th layer, respectively. When employing the activation function R, each layer of discriminator can be represented as

$$L(x) = R\left[\sum_{i=1}^{6} F(x_i, w_i) + x_0\right] = R\left[\sum_{i=1}^{6} w_i x_i + x_0\right] \tag{15}$$

According Lipschitz's inequality, we have

$$\begin{aligned}&||L||_{\mathrm{Lip}} = \\ &||\nabla L(x)||_2 = \left\|R\left[\sum_{i=1}^{6} w_i x_i + x_0\right]\right\|_2 = \\ &||R_6 W_6 (R_5 W_5 (\cdots R_1 W_1(x))) + x_0||_2 \leqslant \\ &||R_6||_2 ||W_6||_2 \cdots ||R_1||_2 ||W_1(x)||_2 + ||x_0||_2\end{aligned} \tag{16}$$

where $W_i(\cdot) = w_i x_i$ represents the weighted sum of the inputs of the i-th layer with its corresponding weight matrix w_i, and the calculated result serves as the input for the next layer. x denotes the first input. $R_6(\cdot)$ to $R_1(\cdot)$ represent the non-linear mapping of the output of each layer using activation functions, respectively. According to the definition of the spectral parametrization[3],

$$\sigma(A) := \max_{||h||\neq 0} \frac{||Ah||_2}{||h||_2} = \max_{||h||=1} ||A||_2 \tag{17}$$

where matrix A can be of any shape, while h is a non-zero vector in the column space of matrix A. $||W||_2$ denotes spectral parametrization of the parameter matrix W and $\sigma(R)=1$, from which it follows that

$$\begin{aligned} &||L||_{\text{Lip}} = \\ &||R_6 W_6(R_5 W_5(\cdots R_1 W_1(x))) + x_0||_2 \leqslant \\ &||R_6||_2 ||W_6||_2 \cdots ||R_1||_2 ||W_1(x)||_2 + ||x_0||_2 \leqslant \\ &\sigma(W_6)\sigma(W_5)\cdots\sigma(W_1)\sigma(x_0) \leqslant \\ &\prod_{i=1}^{6} \sigma(W_i)\sigma(x_0) \end{aligned} \tag{18}$$

where $\sigma(W)$ denotes maximum singular value of matrix W. Note that the input x_0 will be connected with the output of $(i-1)$-th layer, which means that weight matrix has no contribution on x_0. Thus, the Eq. (17) can be written as

$$\begin{aligned} &||L||_{\text{Lip}} = ||\frac{W_n}{\sigma(W_n)} \frac{W_{n-1}}{\sigma(W_{n-1})} \frac{W_{n-2}}{\sigma(W_{n-2})} \cdots \frac{W_1}{\sigma(W_1)}||_2 \leqslant \\ &\prod_{i=1}^{n} \frac{\sigma(W_i)}{\sigma(W_i)} = 1 \end{aligned} \tag{19}$$

From Eq. (19), it can be seen that the discriminator which based on ResBlocks can also satisfy the Lipschitz constraint with $\gamma = 1$ by simply letting parameter matrix of each network layer in discriminator be divided by the spectral norm of the corresponding parameter matrix.

3.3 Analysis of training speed

It can be generally concluded that it is time-saving by training one generator to learn latent shared representations comparing with training multiple generators. We now try to compare training fashion in an accurate way. ADGAN firstly uses three DenseBlock for upscaling in shrinkage path, and each contains two densely connected DenseLayers, i.e.,

$$h_l = F_l([h_{l-1}, h_{l-2}, \ldots, h_0]) \ \ (l = 2,3,4,5,6,7) \tag{20}$$

where h_l represents the output of the l-th layer and F_l denotes nonlinear processing. Equation (20) indicates that channel-level connections are used between DenseLayers, and the feature map output of each DenseBlock is connected to all previous layers. Then, ADGAN uses eight ResBlocks, i.e.,

$$h_l = F_l(h_{l-1}) + h_{l-1} \ \ (l = 8,9,10,11,12,13,14,15) \tag{21}$$

Actually, Eq. (21) contains the information reuse operation in DenseBlock. Thus, it is theoretically possible to maintain low-level features including texture structure and gray information, which are confirmed by subsequent experimental results in Section 4.4. Let B be convolutional kernel size, and M be the number of parameters of convolutional layer. Let C and O be the number of input and output channel, respectively. Then, the number of parameters of the network block can be expressed as

$$M = B \times B \times C \times O \tag{22}$$

As mentioned before, ADGAN uses 3 DenseBlocks and 8 ResBlocks. Then, according to Eq. (22), there will be about 1 247 232 parameters for ADGAN, and about 7 475 200 parameters for StarGAN V2 which uses 12 ResBlocks, i.e., the latter has about 6 times as many parameters as the former. In this view, the trained ADGAN cost less memory than trained StarGAN V2, which is also confirmed by subsequent experimental results in Section 4.3.

4 Experiment and Analysis

Two high-quality datasets with paired multimodal images are utilized to verify the effectiveness of the proposed ADGAN. We first compare our proposed ADGAN with recent state-of-the-art synthesis methods and then present and analyze the qualitative and quantitative results. In addition, we conduct ablation experiments to show the impact of each loss on medical image cross-modal synthesis. The implementation is based on the Python deep learning frameworks PyTorch and OpenCV, and the source code is available at https://github.com/LEI-YRO/ADGAN.

4.1 Dataset

BraTS2018[34] is a commonly used dataset that contains 220 cases of high-grade glioma and 54 cases of low-grade glioma obtained by collecting multicontrast MR images from 285 cases. Each case includes four modalities, including T1, T1ce, T2, and FLAIR. The training, test, and validation sets are randomly divided at a ratio of 7:2:1. Specifically, each modality contains 900 images in the training set, 103 images in the testing set, and 100 images in the validation set.

The Brain dataset[3] is a whole-brain dataset with multimodal brain scan images of the normal population, cerebrovascular disease population, tumor disease population, degenerative disease population, inflammatory or infectious disease population, and multiple modalities and contains multiple modalities, including MR, CT, and PET. We acquired paired slices of more than 6000 pairs. Among them, 143 pairs of CT images and MR-T2 images are selected as the training set, and 70 pairs are selected as the test set. A total of 108 pairs of MR-T1 images and MR-T2 images are used as the training set, and 49 pairs are used as the test set.

4.2 Experimental details and evaluation metrics

All experiments are performed on 20-core Intel(R) Xeon(R) Gold 6148 CPU and 8 Tesla V100-SXM2 GPUs with the same settings. For the well-trained model, the Adam optimizer with $(\beta 1, \beta 2) = (0.9, 0.99)$ is introduced. The initial learning rate is set to 10^{-4}, and the batch size is set to 8. All images are resized to 256×256.

To evaluate the synthesis performance, we choose the BPGAN, Bi-MGAN and StarGAN V2; the first two are bidirectional methods, and the last can be used to achieve multidirectional synthesis. During the evaluation, two metrics, namely, SSIM[35] and MSIM, are used to assess the structural similarity between synthetic and real images. Three additional metrics, namely, Normalized Root Mean Square Error (NRMSE), PSNR[36], and VIF, are utilized to evaluate the visual quality of the synthesized images.

4.3 Training time and memory cost

To analyze the computational cost, we alternately evaluate the training time and memory cost. Specifically, we use the maximum batch size to evaluate the memory cost. The larger the maximum batch size used in the training process is, the lower the memory cost and computational effort are. The comparison results are

Table 2 Comparisons of training time and memory cost.

Method	Training time (h)		Memory cost (image/s)	
	BraTS2018	Brain	BraTS2018	Brain
StarGAN V2	19.4	8.3	8	8
BPGAN	67.8	13.7	48	48
Bi-MGAN	67.8	13.7	48	48
ADGAN	42.7	18.3	8	8

shown in Table 2. Our method incorporates the VGG16 pretraining module, resulting in a slightly longer training time than that of StarGAN V2; however, it is necessary to improve the perceptual details of the synthesized images. In comparison to the BPGAN and Bi-MGAN, our method demonstrates the advantage of training only one generator on the BraTS2018 dataset. Since the BraTS2018 dataset contains four domains with a total of 12 directions, the BPGAN and Bi-MGAN, which are bidirectional synthesis methods, must be trained six times to complete synthesis in all directions. This process not only becomes tedious but also significantly increases the time needed. Clearly, the ADGAN requires only one end-to-end training. A more efficient and streamlined training process not only saves time but also reduces the risk of overfitting and instability.

4.4 Qualitative results

To explore multidirectional and multidomain adaptive synthesis, we first compare the proposed ADGAN with StarGAN V2, which can achieve adaptive generation on the BraTS2018 dataset, which contains four domain images, i.e., FLAIR, T1, T1ce, and T2. The qualitative results of ADGAN and StarGAN V2 on BraTS2018 with paired fashion training are shown in Fig. 3.

As shown in Fig. 3, both our ADGAN and StarGAN V2 can achieve multidirectional synthesis between multiple domains and obtain generally satisfactory synthesized results. However, the synthesized image generated by StarGAN V2 incorrectly reflects grayscale values of the target image, i.e., it does not ensure pathological invariance and misses the lesion information in the source image during the synthesis process. In contrast, the synthesized image generated by the ADGAN maintains correct mapping between the synthesized and source images. The synthesized images retain the shape and lesion information of the midbrain in the source images while being mapped to corresponding images in the reference image domain that are realistic and of high quality.

As mentioned before, the BPGAN and Bi-MGAN can achieve cross-modal bidirectional synthesis. To further validate the synthesis performance of the ADGAN, the BPGAN and Bi-MGAN are compared, and the results are shown in Fig. 4.

BPGANs are mainly used for CT-T2 bidirectional synthesis, and Bi-MGANs are designed for T1-T2 bidirectional synthesis. As shown in Figs. 4a and 4b, the BPGAN performs well when conducting CT-T2 bidirectional synthesis, which is consistent with the results reported for the BPGAN. However, since the BPGAN was originally designed to achieve bidirectional synthesis between CT and T1, the BPGAN fails to achieve bidirectional synthesis between CT and T1. A similar trend can be found in Figs. 4c and 4d. Compared to BPGAN and Bi-MGAN, ADGAN performs almost all domain synthesis tasks. Specifically, compared with BPGANs, T1-T2 bidirectional synthesis has clear advantages and achieves results comparable to those of CT-T2 bidirectional synthesis. Similarly, ADGANs also possess clear advantages for bidirectional CT-T2 synthesis and achieve comparable results for bidirectional T1-T2 synthesis. Compared to BPGAN and Bi-MGAN, ADGAN introduces domain feature reconstruction loss to allow the generator to learn latent shared features across multiple target domains, which is better than constructing a separate nonlinear mapping for different modalities.

Considering all the comparisons in Figs. 3 and 4, it can be concluded that our proposed ADGAN has clear advantages in adaptive and multidirectional cross-modal medical image synthesis. In addition, compared with state-of-the-art bidirectional synthesis methods (i.e., BPGAN and Bi-MGAN), it can also achieve comparable results in specific domains.

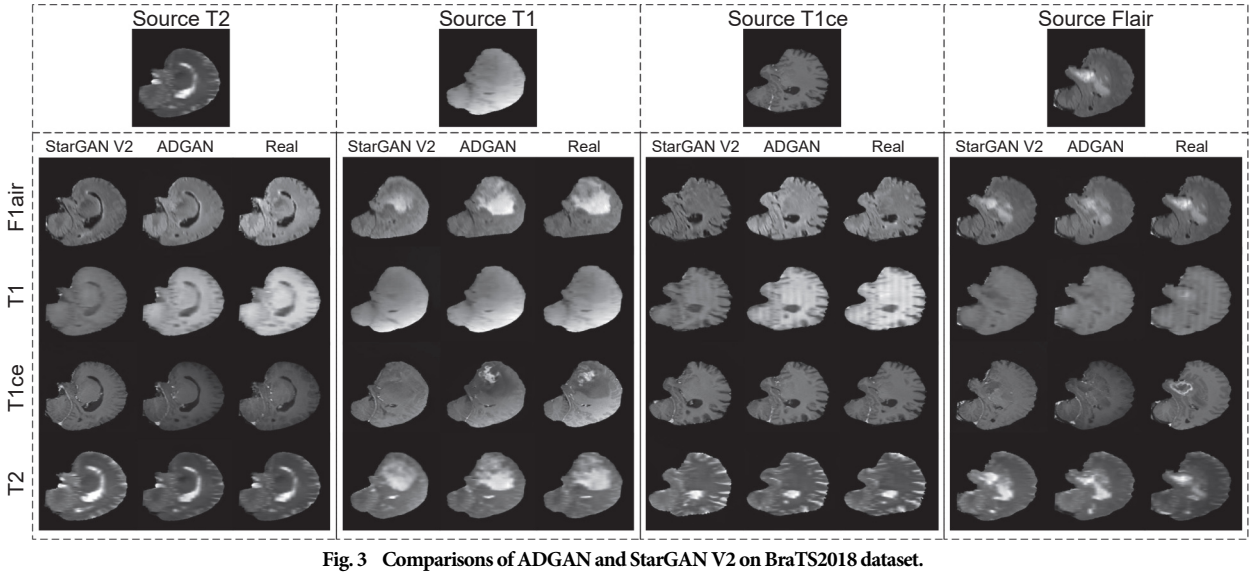

Fig. 3 Comparisons of ADGAN and StarGAN V2 on BraTS2018 dataset.

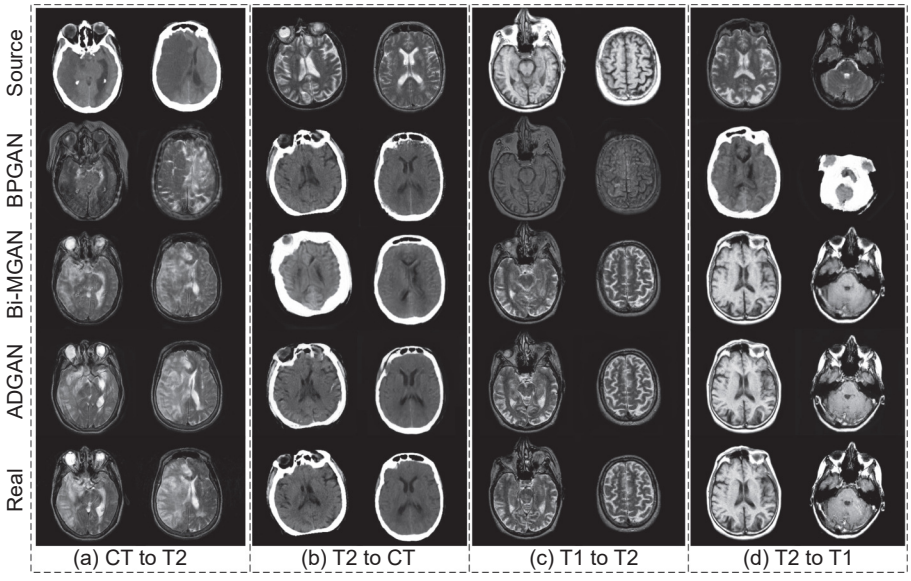

Fig. 4 Comparisons of ADGAN, BPGAN, and Bi-MGAN on Brain.

4.5 Quantitative results

The experimental results in Figs. 3 and 4 demonstrate the competitiveness of the ADGAN compared with the baseline methods. To comprehensively evaluate the proposed ADGAN, we have also obtained quantitative results. The SSIM, MSIM, PSNR, NRMSE, and VIF are used for evaluation, and the results are shown in Tables 3–5.

The synthetic results of ADGAN and StarGAN V2 on the

Table 3 Evaluation metrics of ADGAN and StarGAN V2 on BraTS2018 in 12 directions (↑: Higher is better; ↓: Lower is better).

Method	Direction	SSIM ↑	PSNR ↑	MSIM ↑	NRMSE ↓	VIF ↑
StarGAN V2	Flair→T1	0.715	18.173	0.782	0.416	0.827
	T1→Flair	0.594	18.316	0.731	0.448	0.863
	Flair→T1ce	0.593	21.676	0.749	0.386	0.819
	T1ce→Flair	0.599	18.358	0.723	0.442	0.868
	Flair→T2	0.733	21.748	0.799	0.322	0.816
	T2→Flair	0.611	18.805	0.757	0.425	0.871
	T1→T1ce	0.674	22.758	0.804	0.333	0.814
	T1ce→T1	0.750	19.690	0.841	0.359	0.838
	T1→T2	0.775	22.378	0.807	0.293	0.819
	T2→T1	0.735	19.397	0.821	0.371	0.837
	T1ce→T2	0.770	22.389	0.810	0.291	0.824
	T2→T1ce	0.655	22.471	0.783	0.343	0.819
	Mean	0.684	20.513	0.784	0.369	0.835
ADGAN	Flair→T1	0.836	24.083	0.886	0.183	0.839
	T1→Flair	0.789	24.112	0.862	0.202	0.866
	Flair→T1ce	0.783	25.267	0.847	0.250	0.931
	T1ce→Flair	0.796	23.903	0.864	0.208	0.865
	Flair→T2	0.811	24.910	0.871	0.221	0.838
	T2→Flair	0.799	24.098	0.867	0.203	0.866
	T1→T1ce	0.792	25.758	0.862	0.235	0.833
	T1ce→T1	0.850	24.510	0.906	0.174	0.841
	T1→T2	0.826	25.711	0.889	0.199	0.840
	T2→T1	0.847	24.655	0.903	0.173	0.841
	T1ce→T2	0.828	25.814	0.893	0.198	0.840
	T2→T1ce	0.791	25.704	0.863	0.236	0.829
	Mean	**0.812**	**24.877**	**0.876**	**0.207**	**0.852**

Table 4 Evaluation metrics of the synthetic results of the ADGAN and BPGAN on the Brain dataset.

Method	Direction	SSIM ↑	PSNR ↑	MSIM ↑	NRMSE ↓
BPGAN	CT→T2	0.488	16.663	0.435	0.653
	T2→CT	0.611	12.054	0.523	0.672
ADGAN	CT→T2	**0.653**	**20.825**	**0.739**	**0.440**
	T2→CT	**0.666**	**15.012**	**0.697**	**0.496**

Table 5 Evaluation metrics of the synthetic results of ADGAN and Bi-MGAN on the Brain dataset.

Method	Direction	SSIM ↑	PSNR ↑	MSIM ↑	NRMSE ↓
Bi-MGAN	T1→T2	**0.823**	**24.277**	**0.919**	**0.248**
	T2→T1	**0.846**	**23.521**	**0.935**	**0.152**
ADGAN	T1→T2	0.757	21.510	0.869	0.336
	T2→T1	0.779	19.719	0.887	0.237

BraTS2018 dataset are evaluated in Table 3. As shown in Table 3, we have considered and evaluated 12 synthesis directions among the four image domains. According to Table 3, ADGAN outperforms StarGAN V2 in terms of all the evaluation metrics. Specifically, SSIM, MSIM, PSNR, and VIF achieve average increases of 18.7%, 11.7%, 21.2%, and 2%, respectively. Moreover, the NRMSE decreases by an average of 43.9%. All these results show that the proposed ADGAN outperforms StarGAN V2.

Then, we compare the ADGAN with the bidirectional synthesis methods BPGAN and Bi-MGAN on the Brain dataset, and the results are shown in Tables 4 and 5. From Table 4, it can be seen that the SSIM, MSIM, and PSNR achieve average increases of 2%, 49.8% and 24.8%, respectively. In addition, the NRMSE achieves an average decrease of 29.3%.

As shown in Table 5, the proposed ADGAN does not achieve the highest score in each index and even incurs minor decreases in PSNR and NRMSE compared with the Bi-MGAN on the Brain dataset. Specifically, SSIM and MSIM achieve average increases of 7.9% and 5.2%, respectively. Bi-MGANs incorporate both deep and handcrafted loss during training and yield satisfactory results by considering both high- and low-level semantic information. Despite some advantages possessed by the Bi-MGAN, the proposed ADGAN only contains one generator and one discriminator, which has clear advantages in training time and storage compared with constructing a separate nonlinear mapping for cross-domain synthesis.

Based on all the results in Figs. 3 and 4 and Tables 3–5, it can be concluded that the ADGAN is superior to the StarGAN V2 and BPGAN. Table 5 shows that our ADGAN incurs minor drops compared with the Bi-MGAN in the T1-T2 bidirectional synthesis. However, the ADGAN possesses more flexibility with less training time and computing resources.

4.6 Ablation experiments

The proposed algorithm mainly includes adversarial loss, domain feature reconstruction loss, pixel-level L1 loss, and perceptual loss. To further investigate the effect of the loss function on the synthesis performance, ablation experiments are performed on the ADGAN. ADGAN−D denotes the variant model with domain feature reconstruction loss removed, ADGAN−L denotes the variant model with L1 pixel-level loss removed, and ADGAN−P denotes the variant model with perceptual loss removed; the results are shown in Fig. 5.

As shown in Fig. 5a, all synthesized images have the same shape, which shows that the ADGAN cannot generate synthesized images under the target domain based on the reference image after removing domain feature reconstruction loss and that the model cannot learn the correct mapping relationship. When removing the pixel-level L1 loss, the correct mapping relationship between the source and synthesized images can be maintained, but the edge details of the synthesized images are blurred, and the quality is slightly lower than that of the original method, as shown in Fig. 5b. Additionally, when removing the perceptual loss, Fig. 5c shows that the model can still learn the correct mapping relationship, but the synthesized images show large blurred patches, which means that it can roughly retain the shape of the

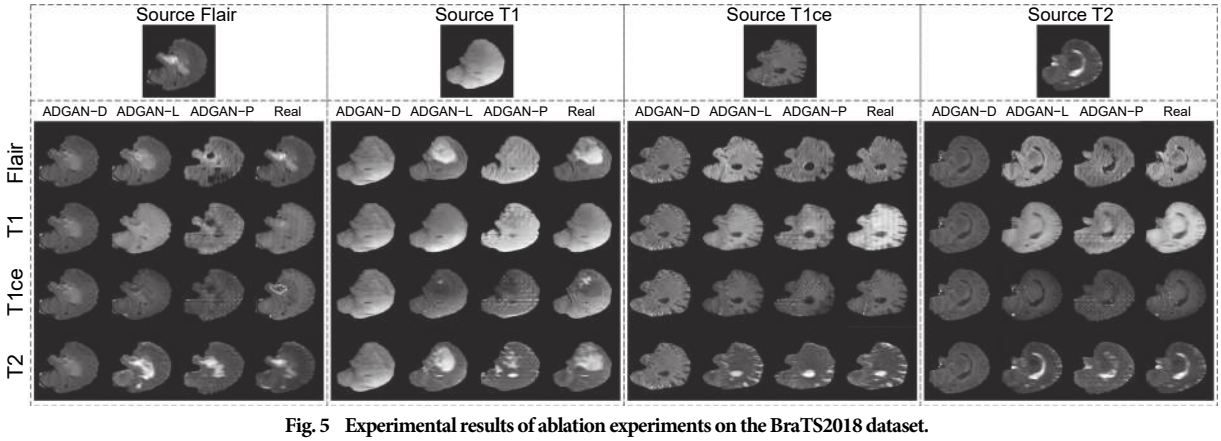

Fig. 5 Experimental results of ablation experiments on the BraTS2018 dataset.

source image but fails to preserve structural details.

To further accurately assess the impact of each loss function on the quality of the synthesized images, the three variant models were objectively evaluated, and the evaluation metrics and results are shown in Table 6.

To simplify the evaluation, we compute the average values of each evaluation metric in the 12 directions shown in Table 3 and list all the evaluations in these directions. As shown in Table 6, the ADGAN obtains the highest score for each index. Compared with ADGAN–D, which removes domain feature reconstruction loss, it achieves the most improvement. As shown in Fig. 5, we can also conclude that removing domain feature reconstruction loss yields worse results.

In order to verify the effect of the hybrid generator, we analyze it with two different structures. By comparing the generators of ResUnet, DenseUnet and hybrid on the same dataset, the results are shown in Table 7. It can be seen that under the same experimental conditions, the hybrid generator performs better than the other two generators.

4.7 Parameter sensitivity analysis

The objective function of the ADGAN contains three hyperparameters (i.e., λ_{per}, λ_{sty}, and λ_{L1}, and we now investigate the effect of each weight. Specifically, we calculate the SSIM by varying the three parameters between 10^{-2} and 10^{3} with a multiplication step of 10. The results on the BraTS2018 and Brain datasets are shown in Fig. 6, from which it can be seen that the maximum value occurs when $\lambda_{L1} = 10^{2}$ and $\lambda_{sty} = \lambda_{per} = 1$. Therefore, we use this combination as the parameter setting scenario.

5 Conclusion

In this paper, we propose an adaptive, multidirectional cross-modal medical image synthesis method. The perceptual loss and pixel-level L1 loss are used to improve the perception and quality of the synthesized image, so that the synthesized image is close to the real image of the target domain and maintains the pathological invariance. The generator is constructed by combining the advantages of residual network and dense connection network, which effectively preserves texture details and semantic features. Furthermore, the spectral normalization is introduced into the discriminator to control its performance to stabilize the model training and indirectly improve the quality of the synthesized image. The experimental results show that the proposed algorithm outperforms recent cross-modal synthesis algorithms for medical images in terms of its subjective and objective evaluation and scalability. The next step will be to explore the shared stream shape among multiple heterogeneous data (e.g., image, audio, and video) to further complete the adaptive and multidirectional synthesis of cross-modal heterogeneous data.

Acknowledgment

The authors would like to thank the anonymous reviewers for their help. This work was supported by the National Natural Science Foundation of China (Nos. 62176217 and 62206224), the Innovation Team Funds of China West Normal University (No.

Table 6 Evaluation of loss function ablation experiments.

Method	SSIM↑	PSNR↑	MSSIM↑	NRMSE↓	VIF↑
ADGAN	**0.812**	**24.877**	**0.876**	**0.207**	**0.852**
ADGAN–D	0.675	18.557	0.681	0.439	0.835
ADGAN–L	0.711	21.253	0.769	0.348	0.837
ADGAN–P	0.677	18.563	0.658	0.452	0.828

Table 7 Evaluation of generator structure ablation experiments.

Generator	SSIM↑	PSNR↑	MSSIM↑	NRMSE↓	VIF↑
ResUnet	0.758	24.351	0.870	0.225	0.843
DenseUnet	0.744	24.104	0.867	0.229	0.845
Hybrid	**0.812**	**24.877**	**0.876**	**0.207**	**0.852**

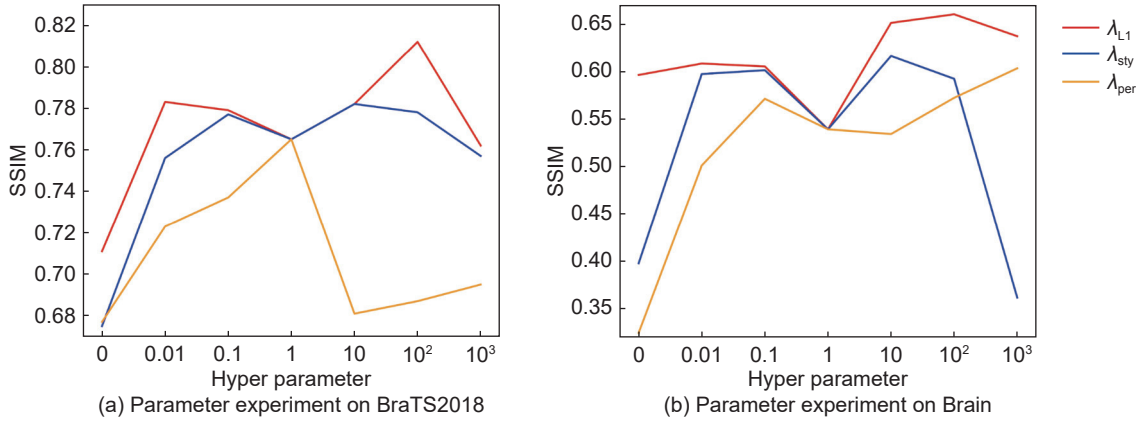

Fig. 6 SSIM results on the BraTS2018 and Brain datasets with hyperparameters in the range of $[10^{-2}, 10^{3}]$.

KCXTD2022-3), the Postdoctoral Science Foundation of China (No. 2023M732428), the Natural Science Foundation of Sichuan Province (No. 2022NSFSC0866), and the Doctoral Research Innovation Project (No. 21E025).

Article History

Received: 8 August 2023; Revised: 28 March 2024; Accepted: 11 April 2024

References

[1] M. J. Willemink, W. A. Koszek, C. Hardell, J. Wu, D. Fleischmann, H. Harvey, L. R. Folio, R. M. Summers, D. L. Rubin, and M. P. Lungren, Preparing medical imaging data for machine learning, *Radiology*, vol. 295, no. 1, pp. 4–15, 2020.

[2] L. Xu, X. Zeng, H. Zhang, W. Li, J. Lei, and Z. Huang, BPGAN: Bidirectional CT-to-MRI prediction using multi-generative multi-adversarial nets with spectral normalization and localization, *Neural Netw.*, vol. 128, pp. 82–96, 2020.

[3] L. Xu, H. Zhang, L. Song, and Y. Lei, Bi-MGAN: Bidirectional T1-to-T2 MRI images prediction using multi-generative multi-adversarial nets, *Biomed. Signal Process. Control*, vol. 78, p. 103994, 2022.

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ImageNet classification with deep convolutional neural networks, *Commun. ACM*, vol. 60, no. 6, pp. 84–90, 2017.

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

[6] M. F. Spadea, G. Pileggi, P. Zaffino, P. Salome, C. Catana, D. Izquierdo-Garcia, F. Amato, and J. Seco, Deep convolution neural network (DCNN) multiplane approach to synthetic CT generation from MR images—Application in brain Proton Therapy, *Int. J. Radiat. Oncol.*, vol. 105, no. 3, pp. 495–503, 2019.

[7] J. Long, E. Shelhamer, and T. Darrell, Fully convolutional networks for semantic segmentation, in *Proc. 2015 IEEE Conf. Computer Vision and Pattern Recognition* (*CVPR*), Boston, MA, USA, 2015, pp. 3431–3440.

[8] A. Ben-Cohen, E. Klang, S. P. Raskin, S. Soffer, S. Ben-Haim, E. Konen, M. M. Amitai, and H. Greenspan, Cross-modality synthesis from CT to PET using FCN and GAN networks for improved automated lesion detection, *Eng. Appl. Artif. Intell.*, vol. 78, pp. 186–194, 2019.

[9] A. Chartsias, T. Joyce, M. V. Giuffrida, and S. A. Tsaftaris, Multimodal MR synthesis via modality-invariant latent representation, *IEEE Trans. Med. Imaging*, vol. 37, no. 3, pp. 803–814, 2018.

[10] G. Li, L. Bai, C. Zhu, E. Wu, and R. Ma, A novel method of synthetic CT generation from MR images based on convolutional neural networks, in *Proc. 2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics* (*CISP-BMEI*), Beijing, China, 2018, pp. 1–5.

[11] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, Generative adversarial networks, *Commun. ACM*, vol. 63, no. 11, pp. 139–144, 2020.

[12] L. Lan, L. You, Z. Zhang, Z. Fan, W. Zhao, N. Zeng, Y. Chen, and X. Zhou, Generative adversarial networks and its applications in biomedical informatics, *Front. Public Health*, vol. 8, p. 164, 2020.

[13] C. Tian, Y. Yuan, S. Zhang, C. W. Lin, W. Zuo, and D. Zhang, Image super-resolution with an enhanced group convolutional neural network, *Neural Netw.*, vol. 153, no. 1, pp. 373–385, 2022.

[14] X. Yang, Y. Lin, Z. Wang, X. Li, and K. T. Cheng, Bi-modality medical image synthesis using semi-supervised sequential generative adversarial networks, *IEEE J. Biomed. Health Inform.*, vol. 24, no. 3, pp. 855–865, 2020.

[15] T. Zhang, H. Fu, Y. Zhao, J. Cheng, M. Guo, Z. Gu, B. Yang, Y. Xiao, S. Gao, and J. Liu, SkrGAN: Sketching-rendering unconditional generative adversarial networks for medical image synthesis, in *Proc. 22nd Int. Conf. Medical Image Computing and Computer Assisted Intervention*, Shenzhen, China, 2019, pp. 777–785.

[16] J. Wang, Y. Zhao, J. H. Noble, and B. M. Dawant, Conditional generative adversarial networks for metal artifact reduction in CT images of the ear, in *Proc. 21st Int. Conf. Medical Image Computing and Computer Assisted Intervention*, Granada, Spain, 2018, pp. 3–11.

[17] M. Yurt, S. U. Dar, A. Erdem, E. Erdem, K. K. Oguz, and T. Çukur, mustGAN: Multi-stream generative adversarial networks for MR image synthesis, *Med. Image Anal.*, vol. 70, p. 101944, 2021.

[18] N. Xiao, X. Xiao, Y. Qiang, K. Li, S. Li, and J. Lian, Longitudinal prediction of lung tumor based on conditional adversarial spatiotemporal encoder (in Chinese), *J. Softw.*, vol. 34, no. 9, pp. 4392–4406, 2023.

[19] B. Yu, L. Zhou, L. Wang, Y. Shi, J. Fripp, and P. Bourgeat, Ea-GANs: Edge-aware generative adversarial networks for cross-modality MR image synthesis, *IEEE Trans. Med. Imaging*, vol. 38, no. 7, pp. 1750–1762, 2019.

[20] Y. Luo, D. Nie, B. Zhan, Z. Li, X. Wu, J. Zhou, Y. Wang, and D. Shen, Edge-preserving MRI image synthesis via adversarial network with iterative multi-scale fusion, *Neurocomputing*, vol. 452, pp. 63–77, 2021.

[21] B. Zhan, D. Li, X. Wu, J. Zhou, and Y. Wang, Multi-modal MRI image synthesis via GAN with multi-scale gate mergence, *IEEE J. Biomed. Health Inform.*, vol. 26, no. 1, pp. 17–26, 2022.

[22] Y. Luo, L. Zhou, B. Zhan, Y. Fei, J. Zhou, Y. Wang, and D. Shen, Adaptive rectification based adversarial network with spectrum constraint for high-quality PET image synthesis, *Med. Image Anal.*, vol. 77, p. 102335, 2022.

[23] S. Ding and J. Lyu, High resolution dermoscopy image synthesis method with pix2pixHD (in Chinese), *J. Comput. -Aided Des. Comput. Graphics*, vol. 32, no. 11, pp. 1795–1803, 2020.

[24] C. Wang, G. Yang, G. Papanastasiou, S. A. Tsaftaris, D. E. Newby, C. Gray, G. Macnaught, and T. J. MacGillivray, DiCyc: GAN-based deformation invariant cross-domain information fusion for medical image synthesis, *Inf. Fusion*, vol. 67, pp. 147–160, 2021.

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

[26] Z. Zhang, L. Yang, and Y. Zheng, Translating and segmenting multimodal medical volumes with cycle- and shape-consistency generative adversarial network, in *Proc. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 2018, pp. 9242–9251.

[27] Y. Pan, M. Liu, C. Lian, T. Zhou, Y. Xia, and D. Shen, Synthesizing missing PET from MRI with cycle-consistent generative adversarial networks for Alzheimer's disease diagnosis, in *Proc. 21st Int. Conf. Medical Image Computing and Computer Assisted Intervention*, Granada, Spain, 2018, pp. 455–463.

[28] G. Zeng and G. Zheng, Hybrid generative adversarial networks for deep MR to CT synthesis using unpaired data, in *Proc. 22nd Int. Conf. Medical Image Computing and Computer Assisted Intervention*, Shenzhen, China, 2019, pp. 759–767.

[29] S. U. Dar, M. Yurt, L. Karacan, A. Erdem, E. Erdem, and T. Cukur, Image synthesis in multi-contrast MRI with conditional generative adversarial networks, *IEEE Trans. Med. Imaging*, vol. 38, no. 10, pp. 2375–2388, 2019.

[30] S. Hu, B. Lei, S. Wang, Y. Wang, Z. Feng, and Y. Shen,

Bidirectional mapping generative adversarial networks for brain MR to PET synthesis, *IEEE Trans. Med. Imaging*, vol. 41, no. 1, pp. 145–157, 2022.

[31] Y. Choi, M. Choi, M. Kim, J. W. Ha, S. Kim, and J. Choo, StarGAN: unified generative adversarial networks for multi-domain image-to-image translation, in Proc. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 2018, pp. 8789–8797.

[32] Y. Choi, Y. Uh, J. Yoo, and J. W. Ha, StarGAN v2: Diverse image synthesis for multiple domains, in *Proc. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Seattle, WA, USA, 2020, pp. 8185–8194.

[33] C. J. Willmott and K. Matsuura, Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance, *Clim. Res.*, vol. 30, pp. 79–82, 2005.

[34] P. Y. Kao, T. Ngo, A. Zhang, J. W. Chen, and B. S. Manjunath, Brain tumor segmentation and tractographic feature extraction from structural MR images for overall survival prediction, in *Proc. 4th Int. Workshop Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, Granada, Spain, 2018, pp. 128–141.

[35] C. Tian, Y. Yuan, S. Zhang, C. W. Lin, W. Zuo, and D. Zhang, Image super-resolution with an enhanced group convolutional neural network, arXiv preprint arXiv: 2204.13620, 2022.

[36] Q. Zhang, J. Xiao, C. Tian, J. Xu, S. Zhang, and C. W. Lin, A parallel and serial denoising network, *Expert Syst. Appl.*, vol. 231, p. 120628, 2023.