Information Fusion 67 (2021) 147–160

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

DiCyc: GAN-based deformation invariant cross-domain information fusion for medical image synthesis

Chengjia Wang [a,*], Guang Yang [b], Giorgos Papanastasiou [c], Sotirios A. Tsaftaris [d], David E. Newby [a], Calum Gray [c], Gillian Macnaught [c,1], Tom J. MacGillivray [e,1]

[a] BHF Centre for Cardiovascular Science, University of Edinburgh, Edinburgh, UK
[b] National Heart and Lung Institute, Imperial College London, London, UK
[c] Edinburgh Imaging Facility QMRI, University of Edinburgh, Edinburgh, UK
[d] Institute for Digital Communications, School of Engineering, University of Edinburgh, Edinburgh, UK
[e] Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK

ARTICLE INFO

Keywords:
Information fusion
GAN
Image synthesis

ABSTRACT

Cycle-consistent generative adversarial network (CycleGAN) has been widely used for cross-domain medical image synthesis tasks particularly due to its ability to deal with unpaired data. However, most CycleGAN-based synthesis methods cannot achieve good alignment between the synthesized images and data from the source domain, even with additional image alignment losses. This is because the CycleGAN generator network can encode the relative deformations and noises associated to different domains. This can be detrimental for the downstream applications that rely on the synthesized images, such as generating pseudo-CT for PET-MR attenuation correction. In this paper, we present a deformation invariant cycle-consistency model that can filter out these domain-specific deformation. The deformation is globally parameterized by thin-plate-spline (TPS), and locally learned by modified deformable convolutional layers. Robustness to domain-specific deformations has been evaluated through experiments on multi-sequence brain MR data and multi-modality abdominal CT and MR data. Experiment results demonstrated that our method can achieve better alignment between the source and target data while maintaining superior image quality of signal compared to several state-of-the-art CycleGAN-based methods.

1. Introduction

Multi-modal medical imaging, i.e. acquiring images of the same organ or structure using different imaging techniques (or modalities) that are based on different physical phenomena, is increasingly used towards improving clinical decision-making. However, collecting data from the same patient using different imaging techniques is often impractical, due to, limited access to different imaging devices, additional time needed for multiple scanning sessions, and the associated cost. This makes cross-domain medical image synthesis a technology that is gaining popularity. We use the term "domain" herein to refer to different imaging modalities, contrast and parametric configurations, for example, for magnetic resonance imaging (MRI). We present a method, called DiCyc, that can perform cross-domain medical image synthesis by learning from non-paired data, thus taking advantage of multiple sources of images, but due to new network architectures it is immune to the presence of deformations inherent to some medical imaging techniques.

Cross-domain image synthesis[2] has been used to impute incomplete information in standard statistical analysis [1,2], to predict and simulate developments of missing information [3], or to improve intermediate steps of analysis such as registration [4], information fusion [5–7], segmentation [8–10], atlas construction [11,12] and disease classification [13,14]. These methods map between MRI, computed tomography (CT), positron emission tomography (PET) and ultrasound imaging from one domain to another. Our main motivation is to synthesize CT images or a particular MR image contrast from multi-sequence MR data. We require the synthesized data should be usable for further medical applications, for example, using synthesized or "pseudo" CT images to improve PET-MR attenuation correction [15–19]. Using MRI to achieve attenuation correction of PET data can be a disadvantage as, unlike CT, the MR signal is not physically related to

* Corresponding author.
E-mail address: chengjia.wang@ed.ac.uk (C. Wang).
[1] These authors contributed equally to this work.
[2] This is also addressed as "image translation" in computer vision.

https://doi.org/10.1016/j.inffus.2020.10.015
Received 4 May 2020; Received in revised form 19 October 2020; Accepted 21 October 2020
Available online 23 October 2020

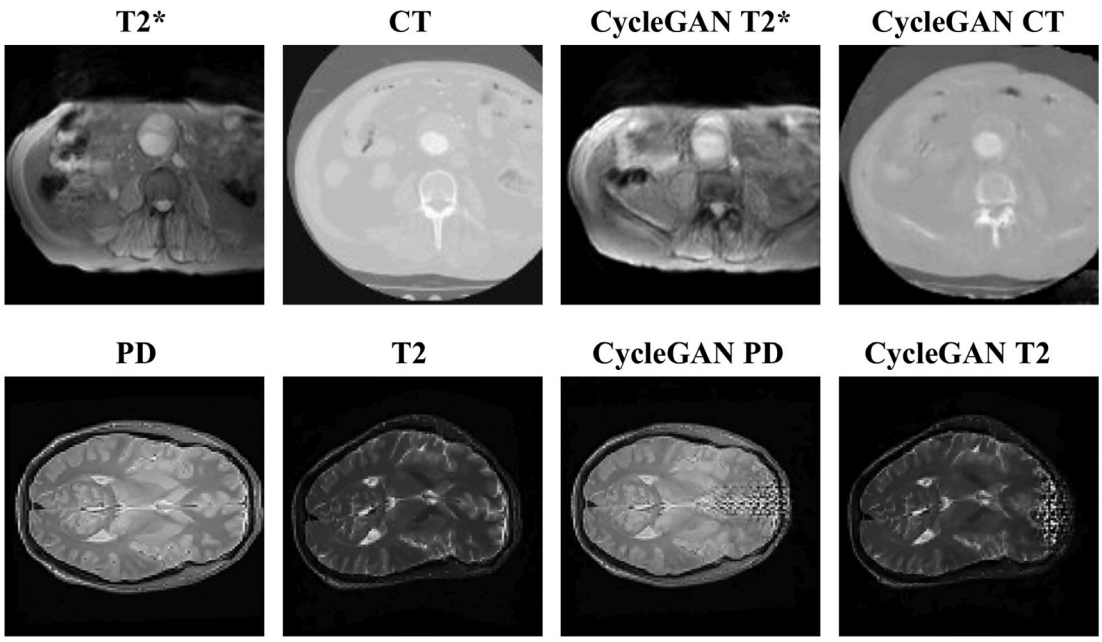

Fig. 1. Example of cross-domain synthesis using vanilla CycleGAN. The first row shows the results obtained from cross-modality abdominal MR-CT data; the second row shows the results of multi-sequence brain data with a synthesized deformation. Both cases demonstrate a reproduction of "domain-specific deformation" in the synthesized output.

attenuation of x-rays in tissue. To overcome this, pseudo-CT generated from corresponding MR could be used to compute a map of linear attenuation coefficients (μ-map) and used for attenuation correction of the PET data acquired on a PET-MRI scanner [20]. This requires mapping of the geometric correspondences between CT and MR.

Learning a contextual correspondence between domains requires not only paired, but well-aligned training data. Such data can be generated by a reliable automatic or manual registration algorithm. As a result, the vast majority of cross-modality image synthesis methods are solely applicable to, or evaluated on brain image data [1,2,4,8,13,16–19,21–29], due to the low geometric variance across different imaging modalities for this particular organ. For other organs, most methods require that the data be aligned by affine transformations or small deformations [3,9,25,30–33]. However, very distinct geometric variances may occur among these data. Nonlinear geometric variances are often associated with different modalities, such as those caused by the shape of imaging bed, the field of view and the axial location planning (captured in Fig. 1). We refer to these as "domain-specific deformations", the presence of which can compromise the quality of the synthesis. This depends on whether the network can learn the mapping sufficiently by being invariant to the presence of deformations (which depends on landing on an ideal local minimum of the loss), or whether pre-processing has removed the deformation due to successful registration (which is not always feasible and cannot deal with large field of view differences).

Methods that allow training with unregistered or unpaired data have recently been proposed [34]. Most state-of-the-art methods use deep convolutional neural networks (CNN) as the image generator within a generative adversarial network (GAN) framework [35]. GAN can represent sharp and even intractable probability densities through a nonparametric approach. It has been widely used in medical image analysis, especially for data augmentation and multi-modality image translations, due to its ability of dealing with domain shift [36]. A popular direction for cross-domain image synthesis is to leverage CycleGAN [37] into the training process. Previous studies have shown that CycleGAN can be trained with unpaired brain data [22,28]. However, CycleGAN can mistakenly encode domain-specific deformations as domain specific features and reproduce the deformations in the synthesized output. Fig. 1 demonstrates two examples. The first row shows a synthesis performed between abdominal CT and T2*-weighted MR, while the second row gives an example of T2-weighted and proton density brain MR with a simulated deformation. In both cases, the deformations specific to the input sources are reproduced by CycleGAN in the output. For applications, such as, attenuation correction where voxel-wise attenuation coefficients are computed, domain-specific deformations should be discarded whilst contextual information relating to the cross-domain appearance of anatomical features and organs is retained.

Recently, several modifications of the vanilla CycleGAN have been proposed, to enhance the alignment between data from the source and target domain using an additional image alignment measure [30,32,38]. However, the additional image alignment loss conflicts with the original loss function in CycleGAN. The synthesized data in which the domain-specific deformations are reproduced will lead to a lower adversarial loss (of the discriminator in GAN). At the same time, the reproduced deformations harm the alignment between the source and the synthesized data, which leads to higher alignment loss. As a result, the synthesized data cannot be aligned to the source data particularly well while maintaining a good quality of signal. To address this issue, we propose the deformation invariant CycleGAN model, or DiCyc. Fig. 2 presents the structural differences between the vanilla CycleGAN and the proposed DiCyc generator networks. We introduce a global transformation model and modified layers of the deformable convolutional network (DCN) into the CycleGAN image generator and propose to the use of a novel image alignment loss based on normalized mutual information (NMI). We evaluate the proposed method using both a publicly available multi-sequence brain MR dataset and our private multi-modality (CT, MR) abdominal dataset. DiCyc displayed better ability to handle disparate imaging domains and to generate synthesized images aligned with the source data whilst keeping comparable quality of the output, compared to state-of-the-art models. Furthermore, the ablation experiment demonstrated that, unlike in the state-of-the-art models, the image alignment loss and the GAN loss were minimized together during training without conflicts in DiCyc.

The main contributions of this paper are as follows:

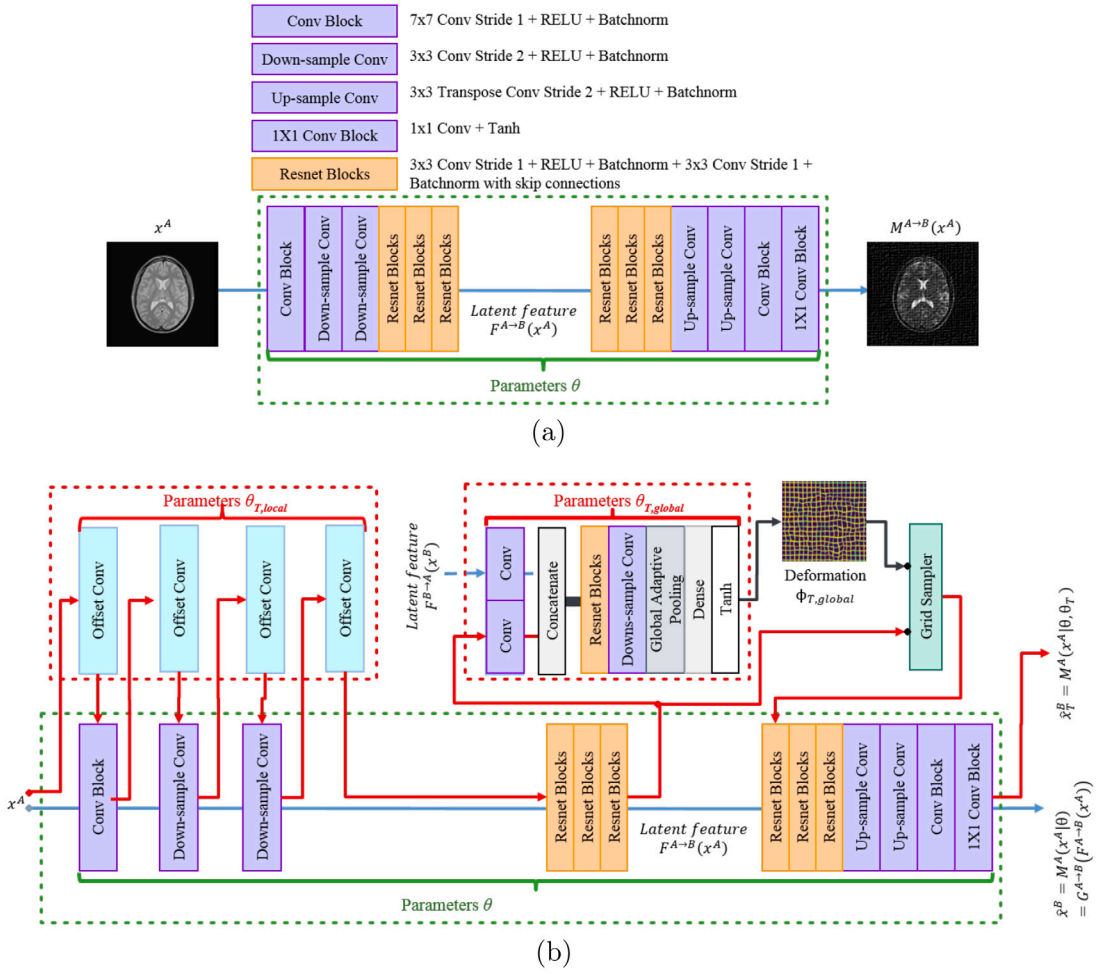

Fig. 2. Comparison of network architectures between CycleGAN and DiCyc. (a) The generator of CycleGAN model used in the original CycleGAN, which is a normal CNN. (b) shows the DiCyc generator network. A deformation convolution layer is inserted in each block before the stack of the Resnet blocks to model the local deformation, parameterized by $\theta_{T,local}$. The global non-linear distortion is modeled using thin-plate-spline (TPS) generated by a spatial transformation subnetwork, parameterized by $\theta_{T,global}$. Details of the modified deformable convolution is shown in Fig. 4(c). The blue arrows represent the CycleGAN forward pass. The additional forward pass introduced by the deformable convolutional layers is represent as red arrows. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

1. We propose a novel DiCyc architecture using a global transformation network and modified deformable convolution layers in between normal convolution layers to address the problem of domain-specific deformations. The deformable layers are modified to have less parameters and offer faster convergence.
2. Rather than the classical "1 forward pass, 1 backward pass" training routine, we designed a new expectation–maximization training procedure where each training iteration includes two distinct forward passes (shown as the blue and red arrows in Fig. 2b) and one single backward pass.
3. We designed a novel cycle-consistency loss and an image alignment loss for information fusion. These losses, together with the new training procedure, address the conflict observed between image alignment loss and the discriminative loss of GAN.
4. We visualized and quantitatively assessed the influence of the domain-specific deformation. We demonstrated the negative effects of the conflict between the image alignment loss and GAN loss in experiments using simulated brain data and realistic abdominal data, and visualized these effects on model convergence in our ablation study.

The paper is organized as follows. Section 2 reviews previous and related techniques. Section 3 gives details of the DiCyc network architecture and the associated loss function. Experiments and datasets used are described in Section 4. The results and discussion are presented in Section 5. Conclusions are given in Section 6.

2. Related works

2.1. Non-CycleGAN models

Typically, most image synthesis methods build up a mapping function from a source to a target domain using paired and pre-registered data. The mapping can be constructed by learning a regression or a dictionary from a collection of patches or feature examples as in [16,17,19,39,40]. Another conventional approach is to build an atlas for each domain using registration, such as modality propagation [41–43]. The prediction is given by mapping between atlases. Along with the rise of deep learning in recent years, neural networks have been used as the cross-domain regressor. For example, the Location Sensitive Deep Network (LSDN) [27] uses a CNN to map the location-dependent patch information between domains. In [26,29], a GAN framework are used to learn the mapping function with context-aware measure based on gradient difference loss. Similarly, [44] uses conditional GAN to synthesize lung histology images. An early method using unpaired data is proposed in [34] where training with unpaired data was addressed as an *unsupervised* approach. It uses mutual information (MI) to select the best corresponding image patches from unpaired cross-domain data, and maximizes a mixture of global MI and local spatial consistency to synthesize multi-sequence brain MR data. This work uses a preprocessing procedure [41] which includes a registration step. Another approach similar to [34] is to construct a dictionary from

patches or image pairs [19,39]. In [45], an algorithm using Weakly-coupled And Geometry (WAG) co-regularized joint dictionary learning is proposed, which learns the patch correspondence from partially unpaired data. Yet, this method was only evaluated using brain images with small geometric variances. A natural strategy in current deep learning based medical image synthesis methods is to model the latent features to arbitrary distributions. For example, [46] assumes the latent features follow a mixed Gaussian distribution, but this method was only evaluated on multi-contrast CT images for segmentation tasks. This paper concentrates on more general synthesis problems between multi-sequence MR data or multi-modal MR and CT data.

2.2. CycleGAN-based methods

CycleGAN was first applied to cross-domain medical image synthesis in [31] and [28] for co-synthesis of CT and MR cardiac and brain data respectively. Both works hint at the influence of deformation affecting results and so removed such artifacts by regularizing the problem through adding additional information (e.g. segmentation masks) [31], and by co-registration [28]. Similarly, in [11,25,30,31,47], the performance of image synthesis networks can be enhanced when jointly trained for segmentation tasks. However, these models require extra manual annotations or registration. Without this requirement, many methods integrate image similarity measures into the GAN loss, for matching the same structure across different domains. For example, [48] introduced a structure-consistency loss based on the modality independent neighborhood descriptor (MIND) [49]. It has been demonstrated that this structure-constrained CycleGAN can deal to some extent with unregistered multi-modal MR and CT brain data. A similar gradient-consistency loss, based on the normalized gradient cross correlation (GCC), is used in [32] for the same purpose. This method has been evaluated using unpaired but pre-registered, multi-modal MR and CT hip images. However, as discussed in Section 1, there is a conflict between the image similarity based losses and the CycleGAN discriminative loss. One potential solution of this problem is to factorize the latent representations into domain-independent semantic features and domain-dependent appearance features, and explicitly filter out the relative spacial deformation between the source and target data [50–52]. This work extends this idea for larger deformations and wider range of domains.

3. Method

3.1. Notation and background

Our goal is to generate synthesized CT or MR data to help post-processing of the source data. For example, a pseudo-CT μmap applicable to PET-MR attenuation correction without registering the synthesized data to the source.

We assume that we have n^A images $x^A \in \mathcal{X}^A$ from domain $\mathcal{X}^A$, and n^B images $x^B \in \mathcal{X}^B$ from domain $\mathcal{X}^B$. For a source image x^A, a generator, $M^{A\to B}$, is trained to generate a synthesized image $\hat{x}^B = M^{A\to B}(x^A)$. Following the GAN setup, $M^{A\to B}$ and a discriminator D^B ares trained to solve the min–max problem of the GAN loss $\mathcal{L}_{GAN}\left(M^{A\to B}, D^B, \mathcal{X}^A, \mathcal{X}^B\right)$. For brevity, we let $\mathcal{L}_{GAN}^{A\to B}$ denote the GAN loss. $M^{A\to B}$ maps the data from $\mathcal{X}^A$ to $\mathcal{X}^B$ while D^B is trained to distinguish whether an image is real or synthesized. Accordingly, for synthesis from $\mathcal{X}^B$ to $\mathcal{X}^A$, there are a $M^{B\to A}$, a D^A, and a GAN loss $\mathcal{L}_{GAN}^{B\to A}$. The vanilla CycleGAN framework consists of two symmetric sets of generators $M^{A\to B}$ and $M^{B\to A}$ act as mapping functions applied to a source domain, and two discriminators D^B and D^A to distinguish real and synthesized data for a target domain [37]. The *cycle consistency* loss $\mathcal{L}_{cyc}\left(M^{A\to B}, D^A, M^{B\to A}, D^B, \mathcal{X}^A, \mathcal{X}^B\right)$, or $\mathcal{L}_{cyc}^{A,B}$, is used to keep the cycle-consistency between the two sets of networks [37]. This gives CycleGAN the ability to deal with unpaired data. Then the loss of the whole CycleGAN framework $\mathcal{L}_{CycleGAN}$ is $\mathcal{L}_{CycleGAN} = \mathcal{L}_{GAN}^{A\to B} + \mathcal{L}_{GAN}^{B\to A} + \lambda_{cyc}\mathcal{L}_{cyc}^{A,B}$. Recent improvements of CycleGAN [32,48] add an image alignment term $\mathcal{L}_{align}^{A,B}$ to $\mathcal{L}_{CycleGAN}$ which becomes

$$\begin{aligned}\mathcal{L}_{CycleGAN,align} &= \mathcal{L}_{CycleGAN} + \lambda_{align}\mathcal{L}_{align}^{A,B} \\ &= \mathcal{L}_{GAN}^{A\to B} + \mathcal{L}_{GAN}^{B\to A} + \lambda_{cyc}\mathcal{L}_{cyc}^{A,B} + \lambda_{align}\mathcal{L}_{align}^{A,B},\end{aligned} \tag{1}$$

where λ_{align} is the weight used to balance the effects of $\mathcal{L}_{align}^{A,B}$ and $\mathcal{L}_{CycleGAN}$. As discussed in Section 1, this causes the conflict between quality of synthesis images and source-target image alignment. The later parts of this section present the detailed analysis of this problem and our DiCyc solution.

3.2. Dicyc architecture

Adding the alignment loss $\mathcal{L}_{align}$ makes cross-domain image synthesis a multi-task learning problem: M is trained for image synthesis while aligning the source and synthesized images. Because the relative deformation, ϕ, between the source and target training images are partially domain specific, this information is encoded by the discriminator D. Note that $\mathcal{L}_{align}$ and $\mathcal{L}_{CycleGAN}$ in existing methods [32,48] are both works on the source image x and the synthesized image $M(x)$. Assuming $M(x)$ is well aligned to x, and $\hat{x}_T^B = M(x)\circ\phi$ is identical to the target image, even when both images have the same image quality, it is always true that

$$\mathcal{L}_{GAN}\left(D^*(x), D^*(M(x))\circ\phi\right) < \mathcal{L}_{GAN}\left(D^*(x), D^*(M(x))\right) \tag{2}$$

for an optimal discriminator D^*. At the same time,

$$\mathcal{L}_{align}(x, M(x)) > \mathcal{L}_{align}(x, M(x)\circ\phi). \tag{3}$$

As a result, $\mathcal{L}_{GAN}$ and $\mathcal{L}_{align}$ lead to gradients with opposite directions: $sgn\left(\nabla_\theta\mathcal{L}_{GAN}\right) \neq sgn\left(\nabla_\theta\mathcal{L}_{align}\right)$ where θ is the network parameters. Any choice of the hyperparameter $\lambda_{align} > 0$ or data augmentation for D will cause a trade-off between the image quality and data alignment.

To solve the problem of inverse gradients, we model the deformation ϕ using a separated set of parameters θ_T. For example, in the $A \to B$ process, $M^{A\to B}$ outputs two synthesized images: one undeformed image aligned to the source:

$$\hat{x}^B = M^{A\to B}(x^A) = M^{A\to B}(x^A|\theta^{A\to B}), \tag{4}$$

and one deformed image that is identical to the target:

$$\hat{x}_T^B = M_T^{A\to B}(x^A) = M^{A\to B}(x^A|\theta^{A\to B}, \theta_T^{A\to B}). \tag{5}$$

As shown in Fig. 1, the relative deformation between the source and target domains can be seen as a combination of a global and a local transformation, thus $\phi = \phi_{global}\circ\phi_{local}$. The corresponding transformation parameters $\theta_T = \{\theta_{T,global}, \theta_{T,local}\}$ are modeled by in different subnetworks in the DiCyc generator (Fig. 2b).

We split the generator M into three subnetworks: an encoder, F, a decoder G and a transformer T. T estimates the global transformation ϕ_{global}, parameterized by $\theta_{T,global}$. In previous CycleGAN based methods parameterize F and F with image synthesize parameters θ. In our DiCyc model, the generator F also estimates the local deformations, parameterized by $\theta_{T,local}$ which is introduced by a series of deformable convolutional layers. As a results, F also produced two versions of latent features: the undeformed feature map $F(x) = F(x|\theta)$ and the locally deformed feature $F_T(x) = F(x)\circ\phi_{local} = F(x|\theta, \theta_{T,local})$.

3.3. Global deformation

The global transformer T has a similar structure with the thin-plate-spline (TPS) based STN. As shown in Fig. 2b, in the $A \to B$ process, the global deformation is calculated by:

$$\phi_{global}^{A\to B} = T^{A\to B}\left(Conv(z^{A\to B}) \oplus Conv(z^{B\to A})\right), \tag{6}$$

where $z^{A\to B}$ and $B \to A$ are latent features given by the encoders $F^{A\to B}$ and $F^{B\to A}$, and $\oplus$ represents the concatenation operation. Specifically,

a regular grid of 6 × 6 control points $t^B = \{t_i^B | i \in \{1, \dots, 36\}\}$ is placed on the latent feature maps of x^B. $T^{A\to B}$ outputs the coordinates of corresponding points t^A on features of x^A. TPS maps the deformation decided by t^A and t^B using an interpolation function Φ. Φ has a form of:

$$t^B = \Phi(t^A) = c + At + W^T s(t), \tag{7}$$

where t is regular image grid and W is the weights assigned to the control points. c and A define the affine transformation between t^A and t^B. s is defined as:

$$s(r) = (\delta(t - t_1), \delta(t - t_2), \dots, \delta(t - t_{36}))^T, \tag{8}$$

where σ is a radial basis kernel has the form of:

$$\delta(r) = r^2 \log r. \tag{9}$$

Note that the transformer T uses a normalized grid where the coordinates $t \in [-1, 1]$.

It has been proved that this form of interpolation function minimizes the bending energy of a surface [53], so it introduces minimal affection on image quality. Based on this analysis, for better quality of synthesis, we wish to keep the local deformation to minimum level within tiny spatial area. When ignoring the local deformation $\phi_{T,local}$, the whole DiCyc model is shown as in Fig. 3.

3.4. Local deformation

We use a modified DCN structure in the encoder G to model the deformation in a local neighborhood after the latent feature $z^{A\to B}$ and $z^{B\to A}$ are globally aligned. A deformable convolutional layer interpolates the input feature maps through an "offset convolution" operation, followed by a normal convolutional layer [54]. This architecture separates the information about local spatial deformation and image context into two forward passes, thus further removes the conflict introduced by $\mathcal{L}_{align}$.

As shown in Fig. 2b, we add an offset convolutional layer (displayed in cyan) before the input convolution layer, the two down-sample convolution layers and the stack of Resnet blocks. This leads to a "lasagne-like" structure consisting of interleaved "offset convolution" and conventional convolution operations so that the spatial deformation is gradually encoded through each layer. The red and blue arrows in Fig. 2b display the computation flows for generating $F_T(x^A)$ and $F(x)$ in the forward passes.

Fig. 4 demonstrates details of the deformable convolution and our modified version used in this work. The deformable convolution can be viewed as an atrous convolution kernel with trainable dilation rates as shown in Fig. 4a. This dilation rate varies across different locations of the input feature maps. As shown in Fig. 4b, the offset of each point in the "N-channel" input feature maps is learned by a standard convolutional operation, outputting 2N "offset maps" (a 2-D deformation for each input feature map is represented by 1 "x" and 1 "y" offset map) [54]. The N input feature maps are then interpolated using the 2N offset feature maps. These operations together are termed as "offset convolution". A standard convolution layer is then applied to the interpolated feature map. When put together these operations form a deformable convolution operation. Designed originally for object recognition tasks, the deformable convolution operation deforms each input feature map independently. Instead, to adjust this operation to cross-domain image synthesis, our modified deformable convolution generates a uniform 2-D deformation that is valid for all input feature maps (Fig. 4c). This is equivalent to directly applying a deformation to the input image and passing it forward through the vanilla CycleGAN generator. This reduces the number of parameters in DCN to a minimum level. Fig. 4d shows our implementation of the "offset convolution".

Combined with the global transformation, $\hat{x}_T^B = F^{A\to B}(F_T^{A\to B}(x^A) \circ \phi_{global}^{A\to B})$ is then taken by the corresponding discriminator D^B to compute GAN losses, and $\hat{x}^B = G^{A\to B}(F^{A\to B}(x^A))$ is expected to be aligned with x^A.

Training DiCyc loss involves the traditional GAN loss, the cycle-consistency loss used in the original implementation of CycleGAN [37], as well as an image alignment loss and an additional cycle consistency loss introduced by the auxiliary outputs obtained from our two separated forward passes. We detail these below.

3.4.1. GAN loss

For the GAN loss $L_{GAN}^{A\to B}$, the minmax game of $M^{A\to B}$ and D^B is represented as:

$$M^{A\to B*}, D^{B*} = \arg\min_{D^B} \max_{M^{A\to B}} \mathcal{L}_{GAN}^{A\to B}, \tag{10}$$

where $M^{A\to B*}$ and D^{B*} represents optimal generator and discriminator. Theoretically, in our DiCyc model, the loss function of D^B is:

$$\mathcal{L}_{GAN}^{D^B} = \mathbb{E}_{x\sim p_{\mathcal{X}^B}} \log\left(D^B(x)\right) + \mathbb{E}_{x\sim p_{\mathcal{X}^A}} \log\left(1 - D^B\left(M_T^{A\to B}(x)\right)\right), \tag{11}$$

where $p_{\mathcal{X}^A}$ and $p_{\mathcal{X}^B}$ represent the data distribution in domain $\mathcal{X}^A$ and $\mathcal{X}^B$. The GAN loss of generator $M^{A\to B}$ is:

$$\mathcal{L}_{GAN}^{A\to B} = \mathbb{E}_{x\sim p_{\mathcal{X}^A}} \log\left(D^B\left(M_T^{A\to B}(x)\right)\right). \tag{12}$$

Similarly, the GAN loss of $M^{B\to A}$ is:

$$\mathcal{L}_{GAN}^{B\to A} = \mathbb{E}_{x\sim p_{\mathcal{X}^B}} \log\left(D^A\left(M_T^{B\to A}(x)\right)\right). \tag{13}$$

3.4.2. Image alignment loss

Eq. (11) can be rewritten as:

$$\mathcal{L}_{GAN}^{D^B} = \mathbb{E}_{x\sim p_{\mathcal{X}^B}} \log\left(D^B(x)\right) + \mathbb{E}_{x\sim p_{\hat{\mathcal{X}}_T^B}} \log\left(1 - D^B\left(M_T^{A\to B}(x)\right)\right), \tag{14}$$

where $p_{\hat{\mathcal{X}}_T^B}$ is the distribution of synthesized domain B images. D^B is then trained to discriminate the distributions $x \sim p_{\mathcal{X}^B}$ and $x \sim p_{\hat{\mathcal{X}}_T^B}$ [35]. In the minmax game of the original GAN model, it has been proved that an optimal discriminator $D^{B*} = p_{\mathcal{X}^B}/(p_{\mathcal{X}^B} + p_{\hat{\mathcal{X}}_T^B})$. Substituting this into $\mathcal{L}_{GAN}^{D^B}$, it can be rewritten as:

$$\begin{aligned}
\mathcal{L}_{GAN}^{D^B} =& \mathbb{E}_{x\sim p_{\mathcal{X}^B}} \log \frac{p_{\mathcal{X}^B}}{p_{\mathcal{X}^B} + p_{\hat{\mathcal{X}}_T^B}} + \mathbb{E}_{x\sim p_{\hat{\mathcal{X}}^B}} \log \frac{p_{\mathcal{X}^B}}{p_{\mathcal{X}^B} + p_{\hat{\mathcal{X}}_T^B}} - \log 4 \\
&+ \mathbb{E}_{x\sim p_{\hat{\mathcal{X}}^B}} \log 4 \\
=& KL\left(p_{\mathcal{X}^B} \,\|\, \frac{p_{\mathcal{X}^B} + p_{\hat{\mathcal{X}}_T^B}}{2}\right) + KL\left(p_{\hat{\mathcal{X}}^B} \,\|\, \frac{p_{\mathcal{X}^B} + p_{\hat{\mathcal{X}}_T^B}}{2}\right) - log4 \\
=& 2 \cdot JSD\left(p_{\mathcal{X}^B} | p_{\hat{\mathcal{X}}_T^B}\right) - log4,
\end{aligned} \tag{15}$$

where KL is the Kullback–Leibler divergence and JSD is the Jensen–Shannon divergence.

Let ψ^A and ψ^B be the spatial poses of the two images, and $\psi^A \sim p_\psi^A$ and $\psi^B \sim p_\psi^B$. For a pair of training images, the relation between ψ^A and ψ^B is:

$$\psi^B = \psi^A \circ \phi^{A\to B} = \psi^A \circ \phi^{-B\to A} = \psi^B \circ \iota, \tag{16}$$

$$\psi^A = \psi^B \circ \phi^{B\to A} = \psi^B \circ \phi^{-A\to B} = \psi^A \circ \iota, \tag{17}$$

where $\phi^{-\cdot}$ represents the inverse transformation and ι represents the identical transformation. With training data which is suffering from the domain-specific deformation, optimally trained D^{B*} and $T^{A\to B*}$ will inevitably predict that $p(x|\psi^A, \phi^{A\to B}) = p_{data}(x^B)$ and $p(x|\psi^A, \iota) \in p_{fake}(x^B)$ even when $\hat{x}^B$ has comparable quality with x^B. As the GAN losses are calculated using $\hat{x}_T^A$ and $\hat{x}_T^B$, a new discriminative loss is required to predict which $p_{\psi,\phi}$ the data is sampled from. Based on the infoGAN theory [55], we can maximize the mutual information (MI) between ϕ and x, as it can be easily proved that

$$\begin{aligned}
MI(x^B, \psi^A | \phi^{A\to B}) =& MI(x^B, \psi^B) = H(x^B) - H(x^B|\psi^B) \\
=& JSD(p_{\mathcal{X}^B} \,\|\, p_{\psi^B}).
\end{aligned} \tag{18}$$

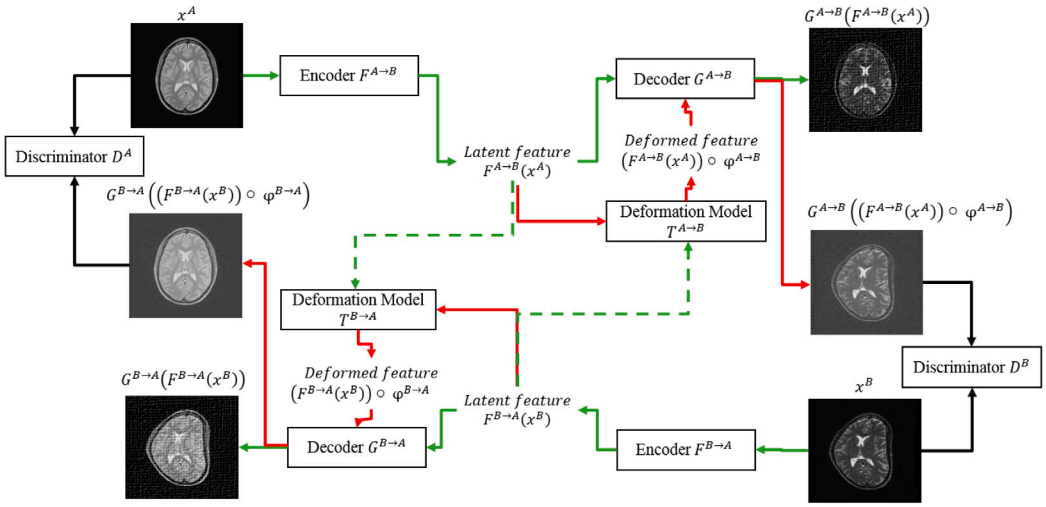

Fig. 3. The DiCyc framework when ignoring local deformation being trained for cross synthesis of PD-weighted (A) and T2-weighted (B) images. The $A \to B$ process is shown by the green arrow and the $B \to A$ process is shown in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

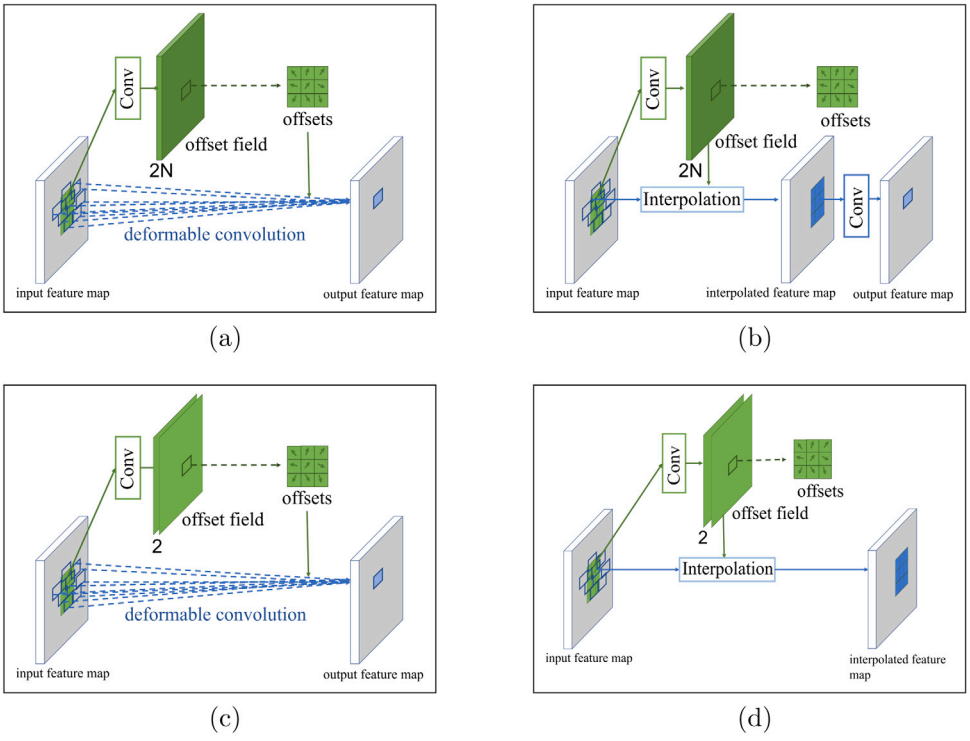

Fig. 4. Details of the original deformable convolution and our modified version. (best viewed in color).

MI yields values from 0 to $+\infty$, which makes it difficult to be scaled and combined with other losses. Here we propose to use an image alignment loss based on NMI:

$$\mathcal{L}_{align}^{A,B} = 2 - NMI\left(x^A, G^{A\to B}\left(x^A\right)\right) - NMI\left(x^B, M^{B\to A}\left(x^B\right)\right). \tag{19}$$

Because the deformations are modeled by a separated set of parameters, this image alignment loss can be adopted with any similarity measure suitable for image registration, such as normalized mutual information (NMI) [56], normalized GCC used in [32], or MIND in [48] and [49].

3.4.3. Cycle-consistency losses

The cycle-consistency loss plays a critical role for the improved performance of CycleGAN compared to a single GAN network, as it forces $M^{A\to B}$ and $M^{B\to A}$ learning mutually recoverable information from distinct domains. As in DiCyc, each generator produces an undeformed and deformed version of synthesized data, both should be cycle-consistent to encode optimal representation. This results in two cycle-consistency losses in our DiCyc model. The undeformed cycle consistency loss is defined as:

$$\mathcal{L}_{cyc}^{A,B} = \| M^{B\to A}\left(M^{A\to B}\left(x^A\right)\right) - x^A \|_1 + \| M^{A\to B}\left(M^{B\to A}\left(x^B\right)\right) - x^B \|_1, \tag{20}$$

and the deformation-invariant cycle consistency loss is:

$$\begin{aligned}\mathcal{L}_{dicyc}^{A,B} =& \| M_T^{B\to A}\left(M_T^{A\to B}\left(x^A\right)\right) - x^A\|_1 \\ &+ \| M_T^{A\to B}\left(M_T^{B\to A}\left(x^B\right)\right) - x^B\|_1 .\end{aligned} \tag{21}$$

3.5. Training procedure

Based on the discussion above, the overall loss of our DiCyc model is[3]

$$\begin{aligned}\mathcal{L}_{DiCyc} = \mathcal{L}_{GAN}^{A\to B} + \mathcal{L}_{GAN}^{B\to A} + \lambda_{align}\mathcal{L}_{align}^{A,B} \\ + \lambda_{cyc}\mathcal{L}_{cyc}^{A,B} + \lambda_{dicyc}\mathcal{L}_{dicyc}^{A,B}.\end{aligned} \tag{22}$$

Treating the cycle-consistent losses as a kind of regularization, training the DiCyc model can be seen as a maximum likelihood estimation (MLE):

$$\begin{aligned}\hat{\theta} &= \arg\max \sum_{(x^A,x^B)} \log p\left(\left(x^A,x^B\right)|\theta\right) \\ &= \arg\max \sum_{(x^A,x^B)} \log \sum_{(\psi^A,\psi^B)} p\left(\left(x^A,x^B\right),\left(\psi^A,\psi^B\right)|\theta\right) \\ &= \arg\max \sum_{(x^A,x^B)} \log \sum_{(\psi^A,\psi^B)} q\left(\left(\psi^A,\psi^B\right)\right) \frac{p\left(\left(x^A,x^B\right),\left(\psi^A,\psi^B\right)|\theta\right)}{q\left(\left(\psi^A,\psi^B\right)\right)}, \\ &= \arg\max \sum_x \log \sum_\psi q(\psi) \frac{p(x,\psi|\theta)}{q(\psi)}\end{aligned} \tag{23}$$

where $q(\psi)$ is an unknown distribution of the image poses. Based on Jensen's inequality, as $\log(\cdot)$ is an convex function,

$$\sum_x \log \sum_\psi q(\psi)\frac{p(x,\psi|\theta)}{q(\psi)} \geq \sum_x \sum_\psi \log q(\psi)\frac{p(x,\psi|\theta)}{q(\psi)}, \tag{24}$$

which gives a lower bound of the maximum likelihood. To make the equality established, $\frac{p(x,\psi|\theta)}{q(\psi)} = c$, where c is a constant. Thus the distribution $q(\psi)$ is:

$$q(\psi) = \frac{p(x,\psi|\theta)}{\sum_\psi p(x,\psi|\theta)} = p(\psi|x,\theta). \tag{25}$$

This MLE learning can be performed through an expectation–maximization (EM) training procedure. The "E" step estimates the distribution $q(\psi)$ by:

$$q\left(\psi_i\right) = q\left(\psi_i|\psi_{i-1},\theta_T^{A\to B},\theta_T^{B\to A}\right), \tag{26}$$

where ψ_{i-1} is decided by the sample training data. For learning optimal global transformations, we fixed the parameters of G and F while only update the STN T. In other world, only the parameters θ_{global} are updated. In the "M" step, the two synthesized images $\hat{x}$ and $\hat{x}_T$ are calculated through two forward passes. The parameters θ are updated based on $\mathcal{L}_{DiCyc}$.

4. Experiments

4.1. Datasets and preprocessing

IXI dataset: We selected two datasets for multi-sequence MR and cross-modality MR-CT data synthesis tasks. The first was the Information eXtraction from Images (IXI) dataset[4] which provides co-registered multi-sequence skull-stripped 1.5T and 3T MR images collected from multiple sites. We used 66 proton density (PD-) and T2-weighted volumes, each volume containing 116 to 130 2D slices. For training and testing, 38 pairs and 28 pairs were used, respectively. Our image generators take 2D axial-plane slices of the volumes as inputs. All volumes were resampled to a resolution of $1.8\times1.8\times1.8$ mm$^3/voxel$, then cropped to a size of 128×128 pixels. As each resampled volume contains 94 to 102 slices, over 6000 pairs of IXI images were used in our experiments. As the generators in both CycleGAN and DiCyc are fully convolutional, the predictions are performed on uncropped images. All the images are bias field corrected and normalized with their mean and standard deviation.

MA^{3}RS dataset: We used a dataset containing 40 pairs of multi-modality abdominal T2*-weighted and CT images collected from 20 patients with abdominal aortic aneurysm. Example images are shown in Fig. 1 where domain-specific deformations can be observed. The data were collected as part of the MA^3RS clinical trial[5] [57]. All images were resampled to a resolution of $1.56\times1.56\times5$ mm$^3/voxel$, and the axial-plane slices trimmed to 192×192 pixels. We used 30 volumes for training and 10 volumes for testing. Each resampled volume contains 24 to 40 slices, which gives over 1200 pairs of slices for our experiments.

4.2. Evaluation metrics

Ideally, alignment between data and the quality of the synthesized images can be evaluated by segmentation-based metrics, such as, Dice index. However, it is difficult to generate the segmentation masks on synthesized data, which can also introduce extra errors in the evaluation. Referring to previous image synthesis works discussed in previous sections, here we use three metrics to evaluate performance of image synthesis: mean squared error (MSE), peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) as typically used by other CycleGAN based methods. Given a volume x^A and a target volume x^B, the MSE is computed as: $\frac{1}{N}\sum_1^N\left(x^B - M^{A\to B}(x^A)\right)^2$, where N is number of voxels in the volume. PSNR is calculated as: $10\log_{10}\frac{\max_B^2}{MSE}$, where $\max_B$ is the maximum voxel value of the image x^B. SSIM is computed as: $\frac{(2\mu_A\mu_B+c_1)(2\delta_{AB}+c2)}{(\mu_A^2+\mu_B^2+c_1)(\delta_A^2+\delta+B^2+c2)}$, where μ and δ^2 are mean and variance of a volume, and δ_{AB} is the covariance between x^A and x^B. c_1 and c_2 are two variables to stabilize the division with weak denominator [58]. Larger PSNR and SSIM, or smaller MSE, indicate a better performance of a synthesis algorithm. These metrics were used to identify the best performing CycleGAN-based method, which we will subsequently refer to as the baseline method. We then evaluated the performance of the proposed DiCyc method compared to this baseline method. A paired t-test was used to the difference in mean MSE, PSNR and SSIM values between DiCyc and selected baseline. For the ablation experiment, a paired t-test was performed on metrics arising from the DiCyc model and its CycleGAN-based counterpart. Differences in performance were considered to be statistically significant when the pvalue resulting from the t-test was less than 0.05.

4.3. Experimental setup

We present three experiments using the two datasets. In the first and second, performance of our DiCyc model was compared to the vanilla CycleGAN [28] and state-of-the-art CycleGAN models with image alignment losses [32,48]. For all experiments, we applied random affine transformations, including translation, rotation, scaling, shearing and flipping, to the input data as augmentations in the training stage,[6] and we manually set that each epoch contains 6000 iterations for better network convergence. After comparing performance of the proposed

[3] Here we set $\lambda_{cyc} = \lambda_{dicyc} = 10$ and $\lambda_{align} = 0.9$.

[4] http://brain-development.org/ixi-dataset/.

[5] http://www.isrctn.com/ISRCTN76413758.

[6] The affine transformation are randomly generated within $[-15, 15]$ translation, $[-15°, 15°]$ rotation, $[0.9, 1.1]$ scaling, $[0.9, 1.1]$ shearing, and random flip with a probability of 0.2. This setup makes sure the affine transformation does not move significant about of the imaged object out of the field of view so that quantitatively assessible results can be obtained.

DiCyc with selected state-of-the-art methods, an ablation study was performed to reveal the influence of DiCyc architecture and learning procedure.

Simulated IXI to identify influence of domain-specific deformation: As the brain organs are mainly rigid structures and rarely suffer from non-linear deformations, ground truth obtained from the registered PD- and T2-weighted image pairs allows evident quantitative assessments. When trained on the registered data, all the methods obtained better performance than when they were trained on unaligned and unpaired data. This provided an upper limit of performance for all the tested methods. To assess the ability of the selected methods to deal with domain-specific deformations, we applied a simulated nonlinear transformation to each T2-weighted image. We performed synthesis experiments using the undeformed PD-weighted images and deformed T2-weighted images to generate undeformed T2-weighted data and deformed PD-weighted data. Minibatches of the input data were sampled from randomly selected patients and slices. When using deformed T2-weighted images to generate synthesized PD data, the ground truth was generated by applying the same nonlinear deformation to the source PD images. Similarly, the ground truth for the synthesized T2-weighted data were the original undeformed T2-weighted data provided in IXI. Values for the three evaluation metrics were computed between the synthesized images and the ground truths. We also qualitatively evaluate the synthesized images using error images as in prior works [26, 29].

MA^{3}RS data: After evaluated on simulated dataset with given ground truths, the methods are further evaluated using realistic data from our **MA^{3}RS** dataset. Due to "domain-specific deformations", the multi-modality images cannot be affinely registered. Specifically, the multiple organs in the pair of images can be hardly aligned at the same time. Furthermore, as non-rigid registration remains an ill-posed problem and lacks a gold standard, we did not non-rigidly register the images to generate ground truth for synthesis. However, several objects, such as aorta and spine, are relatively rigid compared to other surrounding soft tissues such as lower gastrointestinal tract organs. These objects can be separately registered with affine transformations. As a result, performance of synthesis should be assessed by alignment of multiple organs, as well as by quantitative analysis of image quality. In this work, for each volume in the MA^3RS dataset, the anatomy of the aorta was manually segmented (as described in [59]). Multi-modality data acquired from the same patient were affinely registered so that the segmented aortas were well aligned. The manual registration and segmentation were performed by 4 clinical researchers. Signal of the synthesized images was evaluated within the segmentation of aorta using the three metrics described above. Image alignment between the source and synthesized data were visually assessed within both the aorta and spine regions. To sum up, a method with better performance should generate images show better alignment in both the aorta and spine region while achieving lower MSE, higher PSNR, and higher SSIM. In the training stage, the input minibatch was sampled from the same patient but randomly selected slices as described in [48]. The data is augmented with similar transformations that have been applied to the IXI dataset.

Ablated models with different alignment losses: The CycleGAN-based models do not handle the conflict between the additive image alignment losses and the discriminative GAN loss, thus cannot achieve good data alignment without sacrificing quality of the synthesized data. By contrast, the architecture and associated training algorithm of DiCyc handles the geometric deformation and contextual correspondence between the domains separately. This property plays a key role in generating synthesized data that are aligned with source data while maintaining a good performance of contextual synthesis. To prove this argument, it is necessary to analyze the different behaviors of an image alignment loss while being used in CycleGAN and DiCyc frameworks. Furthermore, current CycleGAN-based models use GCC and MIND, but we use a NMI-based alignment loss given in Eq. (19). To verify our proposed alignment loss, it is necessary to compare the performance GCC, MIND and NMI under the same architecture and training procedure.

With these motivations in mind, we performed an ablation experiment using the IXI dataset where different image alignment losses were integrated within both CycleGAN and DiCyc models. Specifically, we replaced NMI-based alignment loss used in the proposed model with the GCC- and MIND-based alignment loss to build a GCC-DiCyc and a MIND-DiCyc. Similarly, our NMI-based alignment loss was added to the CycleGAN loss to build a NMI-CycleGAN. Performance of DiCyc's with different alignment losses were then compared to their CycleGAN-based counterparts. We performed a paired t-test on the evaluation metrics for each pair of CycleGAN and DiCyc models with the same alignment loss to evaluate any improvement in performance introduced by our new architecture. Any improvements introduced by the NMI-based alignment loss can be seen by comparing performance of the DiCyc models using different alignment losses. Evolution of the loss values and synthesis results were also visually assessed throughout the training process.

4.4. Implementation details

We used image generators with 6 Resnet blocks, and 70 × 70 PatchGAN [60] as discriminator networks. Based on the default setup of CycleGAN, we use the LSGAN loss to compute $\mathcal{L}_{GAN}$. Experiments were implemented in PyTorch and paired t-tests were performed using Scipy library. All parameters of, or inherit from, vanilla CycleGAN are taken from the PyTorch implementation of the original paper.[7] The first convolutional layer uses 7×7 kernels, all others use 3×3 kernels. The first convolution output 64 channels of feature maps, followed by layers with 128 and 256 channels. All the convolutions in the Resnet blocks have 256 channels.

For the DiCyc, we set $\lambda_{cyc} = \lambda_{dicyc} = 10$ and $\lambda_{align} = 0.9$. The models were trained with Adam optimizer [61] with a fixed learning rate of 0.0002 for the first 100 epochs, followed by 100 epochs with linearly decreasing learning rate. Here we apply a simple early stop strategy: in the first 100 epochs, when $\mathcal{L}_{DiCyc}$ stops decreasing for 10 epochs, the training will move to the learning rate decaying stage; similarly, this tolerance is set to 20 epochs in the second 100 epochs. For the selected benchmark CycleGAN-based models, unless mentioned above, setup of hyper-parameters follows the original publications. Experiments were performed with nVidia Tesla K80 GPUs provided by the Amazon AWS EC2 cloud computing platform.

5. Results and discussion

This section presents the performance across all models assessed. For each experiment, we visualize the data from the source domain and the synthesized results. Quantitative results are shown in terms of MSE, PSNR and SSIM.

5.1. DiCyc versus CycleGAN-based models on IXI

Fig. 5 shows an example of the synthesized images generated by the methods we tested, along with the error images calculated between the synthesized data and corresponding ground truth. For a fair visual comparison, here we present the results obtained by all the compared baselines with the same non-linear deformation. As the simulated "domain-specific deformation" were applied to the T2-weighted data, the synthesized PD-weighted data should display the same deformation aligned with the source data. Similarly, the synthesized T2-weighted data should be aligned with the source PD-weighted data without showing the simulated deformation. However, as shown in Fig. 5,

[7] https://github.com/junyanz/pytorch-CycleGAN-and-pix2pix.

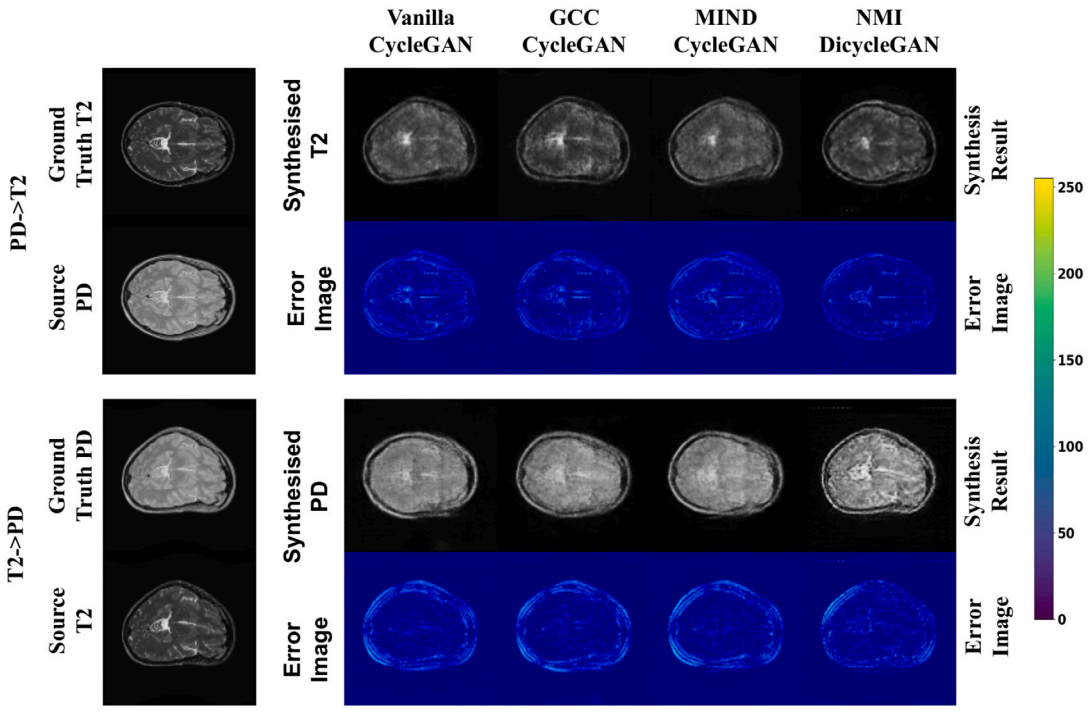

Fig. 5. Examples of synthesis from the IXI dataset: an arbitrary deformation was applied to the T2 weighted images, and the ground truth of the synthesized proton density (PD) weighted image was generated by applying the same deformation.

the vanilla CycleGAN model reproduced the simulated deformation in the synthesized T2-weighted image and did not show the simulated deformation in the synthesized PD-weighted image. Although the GCC-CycleGAN and MIND-CycleGAN reduce the misalignment effect of the simulated deformation, the synthesized and source data are still not well aligned. Furthermore, the synthesis results generated by the three CycleGAN-based models are blurry and showed visible artifacts. In contrast, our DiCyc model gave the best alignment between the source and synthesized data and also lead to better image quality when assessed visually.

The quantitative evaluation of multi-sequence MR synthesis using the IXI dataset is shown in Table 1, where the best result for each metric is shown in bold and the optimum baseline method we chose for a paired t-test is highlighted by a gray background. Vanilla CycleGAN trained on paired and registered images (without simulated deformation) gave the best results with PSNR $>$ 24.3, SSIM $>$ 0.817 and MSE $\leq$ 0.036. This is considered as the upper bound of synthesis performance. Trained with unpaired data that have simulated deformations, the vanilla CycleGAN gave a lower-bound baseline of performance. With additive image alignment losses, GCC-CycleGAN and MIND-CycleGAN methods lead to improvements in terms of PSNR. However, because these two models are still affected by the simulated domain-specific deformation, their performance was still comparable to vanilla CycleGAN.

In contrast, the proposed DiCyc model led to at least 18% increase in MSE, and 8% and 12% performance gain in terms of PSNR and SSIM on IXI data. The results were statistically significant based on the paired t-tests (p-value $<$ 0.05).

5.2. DiCyc versus CycleGAN-based models on MA^{3}RS

Table 2 shows the quantitative assessments of the four models based on the same metrics used for the IXI data. The vanilla CycleGAN had slightly better performance compared to the GCC- and MIND-CycleGAN models. The only exception is that MNID-CycleGAN model obtained higher PSNR in the "T2*→CT" synthesis. Our DiCyc model outperformed the other three methods according to all the metrics.

Table 1
Synthesis results of IXI dataset using deformed T2 images given by value of each metric. Standard deviations are shown within parentheses.

Direction of synthesis: **T2 → PD**			
Method	MSE	PSNR	SSIM
Cycle [28]	0.055 (0.22)	20.80 (2.87)	0.708 (0.19)
GCC-Cycle [32]	0.054 (0.22)	21.04 (3.83)	0.719 (0.19)
MIND-Cycle [48]	0.054 (0.21)	20.82 (2.61)	0.703 (0.19)
Dicycle	**0.045 (0.21)***	**22.52 (2.91)***	**0.790 (0.18)***
Cycle (aligned)	0.037 (0.22)	24.77 (3.30)	0.856 (0.17)
Direction of synthesis: **PD → T2**			
Method	MSE	PSNR	SSIM
Cycle [28]	0.067 (0.19)	18.59 (2.41)	0.671 (0.17)
GCC-Cycle [32]	0.067 (0.20)	18.90 (3.12)	0.684 (0.19)
MIND-Cycle [48]	0.068 (0.20)	18.71 (2.79)	0.687 (0.18)
Dicycle	**0.054 (0.19)***	**20.38 (2.59)***	**0.744 (0.16)***
Cycle (aligned)	0.036 (0.21)	24.30 (3.34)	0.817 (0.18)

*p-value $<$ 0.05.

Note that in the "CT→T2*" synthesis, DiCyc lead to a 20% performance gain in terms of MSE, and achieved 22.8% higher SSIM. Differences between performance achieved by the DiCyc model and the best baseline methods were statistically significant.

The quantitative results shown in Table 2 can be affected by both the qualities of the synthesized images and the alignment between the source and synthesized data. As discussed above, some objects in the images can be affinely registered independently, for example, the anatomy of aorta and spine. However, these two objects cannot be affinely aligned at the same time as a result of domain-specific deformations. This leads to lower PSNR and SSIM, and higher MSE value within the segmented region of aorta.

For better assessing the effects of the domain-specific deformation, the synthesis results of the compared baselines and our TPS-based DiCyc model are displayed in Fig. 6 using a checkerboard visualization. As shown in Fig. 6, when the region of aorta is affinely aligned, the

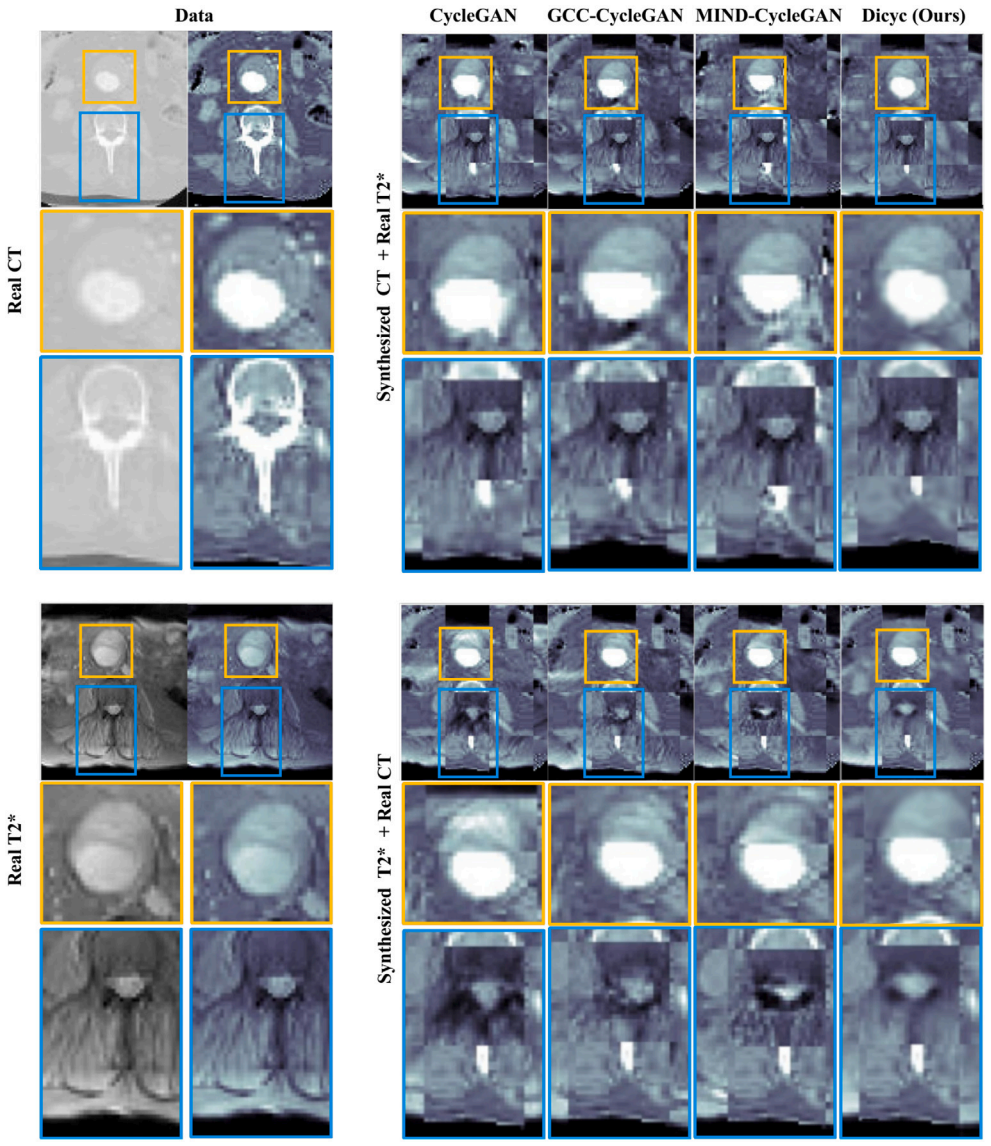

Fig. 6. Visualization results on MA3RS data: the source and the associated synthesized images are displayed using a chessboard visualization. The regions of aorta and spine are highlighted by yellow and blue boxes. CycleGAN-based methods tend to reproduce the domain-specific deformation or suffer from significant artifacts. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Multi-modality synthesis results using private dataset given by value of each metric. Standard deviations are shown within parentheses.

Direction of synthesis: **T2* → CT**			
Model	MSE	PSNR	SSIM
Cycle [28]	0.009 (0.004)	20.57 (2.12)	0.675 (0.06)
GCC-Cycle [32]	0.012 (0.006)	20.25 (2.35)	0.602 (0.08)
MIND-Cycle [48]	0.010 (0.004)	21.21 (2.04)	0.660 (0.07)
Dicycle	**0.008 (0.004)***	**22.01 (2.40)***	**0.694 (0.06)***
Direction of synthesis: **CT → T2***			
Model	MSE	PSNR	SSIM
Cycle [28]	0.025 (0.015)	18.96 (1.36)	0.446 (0.10)
GCC-Cycle [32]	0.040 (0.014)	17.49 (1.20)	0.302 (0.08)
MIND-Cycle [48]	0.034 (0.020)	18.04 (1.49)	0.396 (0.11)
Dicycle	**0.020 (0.014)***	**19.69 (1.35)***	**0.548 (0.12)***

*p-value < 0.05.

CycleGAN-based methods either achieved worse alignment in the spine area, for example, the synthesized CT produced by CycleGAN and GCC-CycleGAN, and the synthesized T2* weighted image given by GCC-CycleGAN; or they generated significant artifacts, for example, in the aorta area of synthesized CT output by CycleGAN and MIND-CycleGAN. Our DiCyc model is the only model that produces synthesized images where both the aorta and spine are simultaneously aligned. Although the synthesized T2* weighted images looks slightly blurred, our DiCyc model generated less artifacts.

5.3. Ablation study

Fig. 7 presents the synthesized images produced by the ablated models using different alignment losses, and the quantitative evaluation results are shown in Table 3. As shown in Fig. 7, all the DiCyc-based models achieved better alignment between the source and synthesized data. This is consistent with the quantitative results shown in Table 3 where in most cases ablated DiCyc models achieved lower MSE and higher PSNR and SSIM models. However, using GCC- and MIND-based alignment losses within the DiCyc framework caused a shift of intensities in the synthesized data. The most obvious example is the synthesized T2-weighted image produced by MIND-DiCyc which

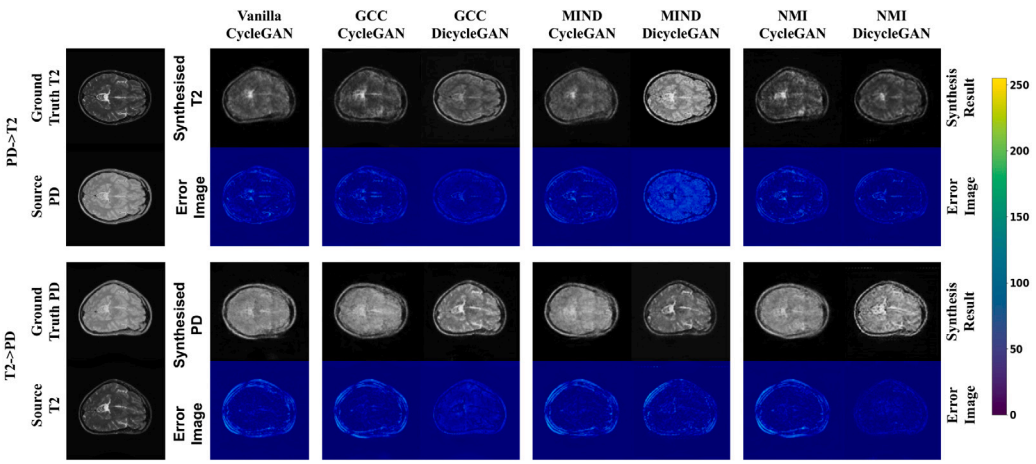

Fig. 7. Visualization of results of ablated models with different image alignment losses. The results were obtained from the IXI dataset with the same simulated deformation applied to the PD-weighted MRI data. The difference image for each method is shown under the synthesis result it generated.

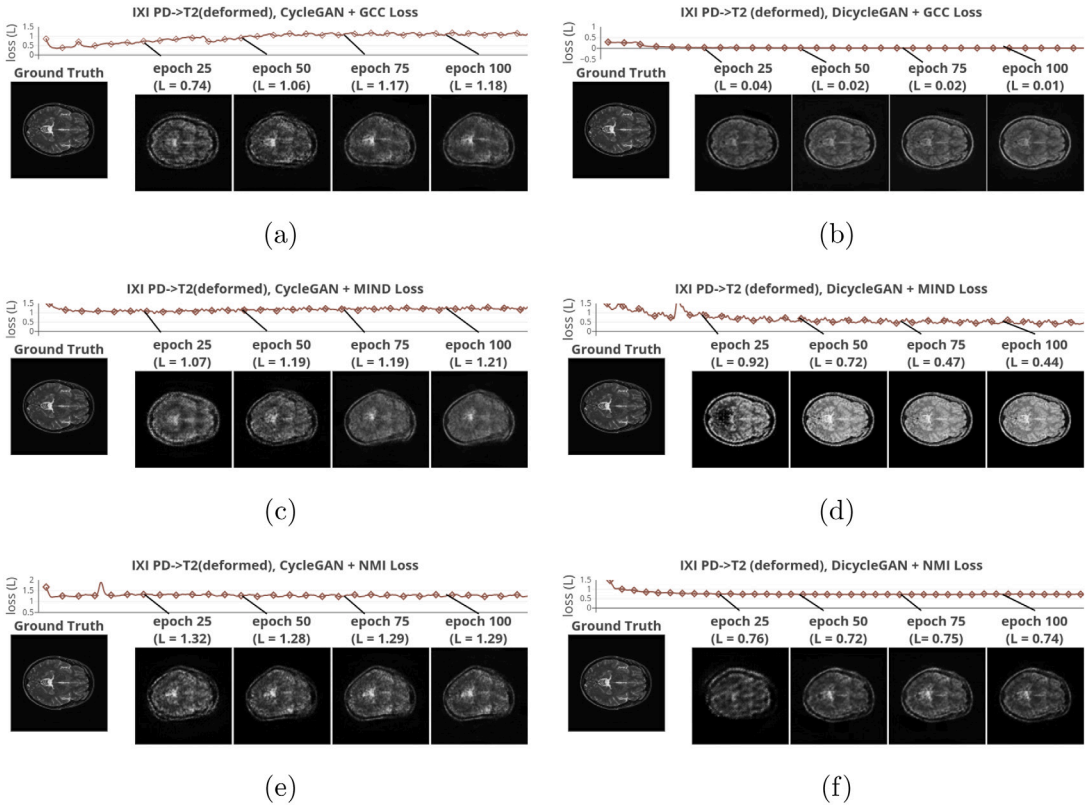

Fig. 8. Evolution of synthesized data during the training process. 8a to 8f successively display the loss curves of GCC-CycleGAN [32], GCC-DiCyc, MIND-CycleGAN [48], MIND-DiCyc, NMI-CycleGAN and NMI-DiCyc (proposed). synthesized T2 weighted data obtained at the 25th, 50th, 75th, 100th epochs are shown above the curves, in comparison of the ground truth shown at the bottom right. (Best viewed in color).

looks more like the source PD-weighted data rather than the target T2-weighted data. As a result, the MIND-DiCyc model gave higher MSE and lower PSNR values in the "PD→T2" synthesis. By contrast, this intensity shift was not observed in the synthesized data generated by our proposed NMI-based DiCyc model. The proposed NMI-DiCyc model outperformed the ablated GCC-DiCyc and MIND-DiCyc models, as well as the state-of-the-art CycleGAN-based methods.

Figs. 8 and 9 demonstrate the evolution of the compared image alignment losses and the synthesis results in the CycleGAN and DiCyc frameworks during the training process. Comparing the synthesis results produced by CycleGAN-based methods (Figs. 8a, 8c and 8e) with those generated by DiCyc models (Figs. 8b, 8d and 8f), we can see that the CycleGAN methods can achieve a good data alignment within the first 20 epochs of training. However as the training algorithm continues to minimize the CycleGAN losses, the domain-specific deformation is gradually reproduced. As the DiCyc framework separately trains the image alignment loss and the CycleGAN loss in two forward passes, the relative deformation between the source and target domain is removed. As shown in Figs. 9a, 9b and 9c, in the CycleGAN framework, each alignment loss was minimized at a certain point of the training process, but then kept increasing as it started to conflict with the GAN discriminative losses. In our DiCyc framework, the alignment losses kept decreasing throughout the whole training process.

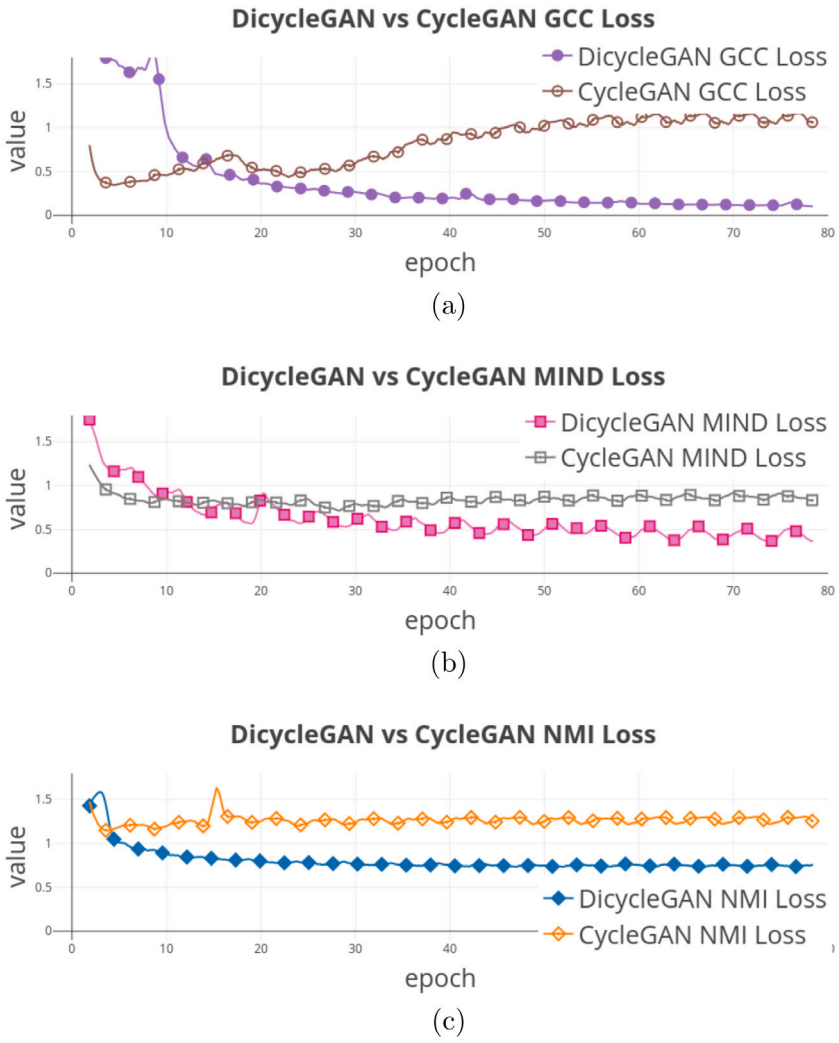

Fig. 9. Evolution processes of each alignment loss in the training process while used in the CycleGAN and DiCyc framework: 9a GCC Loss, 9b MIND Loss, and 9c NMI Loss. (best viewed in color).

Table 3
Multi-modality synthesis results using private dataset given by value of each metric. Standard deviations are shown within parentheses.

Direction of synthesis: **PD → T2**			
Model	MSE	PSNR	SSIM
GCC-Cycle [32]	0.054 (0.021)	21.04 (3.83)	0.719 (0.19)
GCC-Dicycle	**0.047 (0.006)***	**22.17 (3.05)***	**0.840 (0.18)***
MIND-Cycle [48]	**0.054 (0.21)**	**20.82 (2.61)**	0.703 (0.19)
MIND-Dicycle	0.090 (0.22)*	18.59 (2.04)*	**0.714 (0.20)***
NMI-Cycle	0.055 (0.22)	21.03 (3.06)	0.712 (0.20)
Dicycle (NMI)	**0.045 (0.21)***	**22.52 (2.91)***	**0.790 (0.18)***
Direction of synthesis: **T2 → PD**			
Model	MSE	PSNR	SSIM
GCC-Cycle [32]	0.067 (0.20)	18.90 (3.12)	0.684 (0.19)
GCC-Dicycle	**0.054 (0.20)***	**20.42 (3.18)***	**0.740 (0.20)***
MIND-Cycle [48]	0.068 (0.20)	18.71 (2.79)	0.687 (0.18)
MIND-Dicycle	**0.054 (0.19)***	**20.33 (2.97)***	**0.740 (0.19)***
NMI-Cycle	0.067 (0.21)	18.76 (3.09)	0.684 (0.19)
Dicycle (NMI)	**0.054 (0.19)***	**20.38 (2.59)***	**0.744 (0.19)***

*p-value < 0.05.

Comparing the results shown in Figs. 8b, 8d and 8f, we can see that the ablated GCC- and MIND-DiCyc models reproduced the appearance of the PD-weighted data in the synthesized T2-weighted data. This means GCC and MIND are still more domain-dependent measures compared to NMI although they have been widely used in multi-modality registration methods. However, computationally GCC and MIND are easily vectorized and the associated backward pass are easier to implement with lesser computational complexities.

5.4. Model complexity

For the CycleGAN-based baselines compared above, each generator network, M, has 34.52M trainable parameters, and each descriminator network, D, has 2.76M. As a result, in the training stage, a CycleGAN-based model has 74.56M trainable parameters and each forward pass consists of 37.98G multiply-add operations (MACs)[8] processing 128 × 128 image data. For our DiCyc model, the local and the

[8] 1G multiply-add operation (MACs) is roughly 2G floating points operations (FLOPs). Results are obtained using the ptflops package at https://github.com/sovrasov/flops-counter.pytorch and the torchsummary package at https://github.com/sksq96/pytorch-summary.

global transformation modules introduce 8.15M and 4.31M trainable parameters. Each forward pass consists of 66.36G MACs. As a result, it takes 75% more time and 33% extra memory to train a DiCyc model. However, once trained, prediction of the synthesized images is performed only by the image generator without global and local deformation modules. In other word, in the testing stage, the proposed DiCyc model has the same temporal and spacial complexity with the CycleGAN-based methods (34.52M trained parameters, 18.24G MACs per forward pass).

6. Conclusion

We introduced the DiCyc cross-domain medical image synthesis model which addresses the issue of and is resilient to domain-specific deformations. We integrated a modified deformable convolutional layer into the network architecture, and proposed the associated deformation-invariant cycle consistency loss and NMI-based alignment loss function. Experiments were performed for synthesis of multi-sequence MRI data with simulated deformations and of multi-modality CT and MRI data suffering from actual domain-specific deformations. We compared our method to the vanilla CycleGAN method and two state-of-the-art methods with additional alignment losses. Our DiCyc method achieved better alignment between the source and synthesized data while maintaining signal qualities of the synthesized data. It outperformed state-of-the-art methods. In order to reveal the mechanism of DiCyc that is separately encoding the information about spatial deformation in the synthesis process, we also performed an ablation study by integrating popular image similarity metrics into DiCyc and comparing their CycleGAN-based counterparts. It shows that the DiCyc model avoids the conflict between the CycleGAN loss and the image alignment losses. Our NMI-based image alignment loss also demonstrated better robustness for synthesis of images from different domains.

CRediT authorship contribution statement

Chengjia Wang: Conceptualization, Methodology, Software, Data curation, Writing - original draft, Visualization, Investigation, Writing - review & editing. **Guang Yang:** Conceptualization, Investigation, Visualization, Investigation, Writing - review & editing, Validation. **Giorgos Papanastasiou:** Methodology, Data curation, Writing - review & editing, Software, Investigation, Validation. **Sotirios A. Tsaftaris:** Methodology, Validation, Writing - review & editing. **David E. Newby:** Supervision, Validation. **Calum Gray:** Software, Data curation, Visualization, Methodology. **Gillian Macnaught:** Conceptualization, Data curation, Validation, Software, Writing - review & editing, Supervision. **Tom J. MacGillivray:** Conceptualization, Investigation, Validation, Software, Writing - review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is funded by British Heart Foundation, UK (no. RG/16/10 /32375). D.E. Newby is supported by the British Heart Foundation (CH/09/002, RG/16/10/32375, RE/18/5/34216) and is the recipient of a Wellcome Trust Senior Investigator Award (WT103782AIA). G. Yang is supported by IIAT Hangzhou, the British Heart Foundation (Grant Number: PG/16/78/32402), the European Research Council Innovative Medicines Initiative on Development of Therapeutics and Diagnostics Combatting Coronavirus Infections Award (H2020-JTI-IMI2 101005122), and the AI for Health Imaging Award (H2020-SC1-FA-DTS-2019-1 952172). S.A. Tsaftaris and G. Papanastasiou acknowledge support from the EPSRC, UK Grant (EP/P022928/1). S.A. Tsaftaris acknowledges the support of the Royal Academy of Engineering, UK and the Research Chairs and Senior Research Fellowships scheme.

References

[1] G. van Tulder, M. de Bruijne, Why does synthesized data improve multi-sequence classification? in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 531–538.
[2] A.V. Dalca, K.L. Bouman, W.T. Freeman, N.S. Rost, M.R. Sabuncu, P. Golland, Medical image imputation from image collections, IEEE Trans. Med. Imaging (2018).
[3] K. Eilertsen, L. Nilsen Tor Arne Vestad, O. Geier, A. Skretting, A simulation of MRI based dose calculations on the basis of radiotherapy planning CT images, Acta Oncol. 47 (2008) 1294–1302.
[4] J.E. Iglesias, E. Konukoglu, D. Zikic, B. Glocker, K. Van Leemput, B. Fischl, Is synthesizing MRI contrast useful for inter-modality analysis? in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2013, pp. 631–638.
[5] J. Du, W. Li, K. Lu, B. Xiao, An overview of multi-modal medical image fusion, Neurocomputing 215 (2016) 3–20.
[6] Q. He, X. Li, D.N. Kim, X. Jia, X. Gu, X. Zhen, L. Zhou, Feasibility study of a multi-criteria decision-making based hierarchical model for multi-modality feature and multi-classifier fusion: Applications in medical prognosis prediction, Inf. Fusion 55 (2020) 207–219.
[7] K. Wang, M. Zheng, H. Wei, G. Qi, Y. Li, Multi-modality medical image fusion using convolutional neural network and contrast pyramid, Sensors 20 (2020) 2169.
[8] S. Roy, A. Carass, J. Prince, A compressed sensing approach for MR tissue contrast synthesis, in: Biennial International Conference on Information Processing in Medical Imaging, Springer, 2011, pp. 371–383.
[9] A. Chartsias, T. Joyce, G. Papanastasiou, S. Semple, M. Williams, D. Newby, R. Dharmakumar, S.A. Tsaftaris, Factorised spatial representation learning: application in semi-supervised myocardial segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2018, pp. 490–498.
[10] L. Li, X. Zhao, W. Lu, S. Tan, Deep learning for variational multimodality umor segmentation in pet/ct, Neurocomputing 392 (2020) 277–295.
[11] N. Cordier, H. Delingette, M. Lê, N. Ayache, Extended modality propagation: image synthesis of pathological cases, IEEE Trans. Med. Imaging 35 (2016) 2598–2608.
[12] O. Commowick, S.K. Warfield, G. Malandain, Using frankensteins creature paradigm to build a patient specific atlas, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2009, pp. 993–1000.
[13] R. Li, W. Zhang, H.-I. Suk, L. Wang, J. Li, D. Shen, S. Ji, Deep learning based imaging data completion for improved brain disease diagnosis, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2014, pp. 305–312.
[14] T. Zhou, K.-H. Thung, M. Liu, F. Shi, C. Zhang, D. Shen, Multi-modal latent space inducing ensemble svm classifier for early dementia diagnosis with neuroimaging data, Med. Image Anal. 60 (2020) 101630.
[15] G. Wagenknecht, H.-J. Kaiser, F.M. Mottaghy, H. Herzog, MRI for attenuation correction in pet: methods and challenges, Magn. Reson. Mater. Phys. Biol. Med. 26 (2013) 99–113.
[16] A. Torrado-Carvajal, J.L. Herraiz, E. Alcain, A.S. Montemayor, L. Garcia-Cañamaque, J.A. Hernandez-Tamames, Y. Rozenholc, N. Malpica, Fast patch-based pseudo-CT synthesis from T1-weighted MR images for PET/MR attenuation correction in brain studies, J. Nucl. Med. 57 (2016) 136–143.
[17] N. Burgos, M.J. Cardoso, K. Thielemans, M. Modat, S. Pedemonte, J. Dickson, A. Barnes, R. Ahmed, C.J. Mahoney, J.M. Schott, et al., Attenuation correction synthesis for hybrid PET-MR scanners: application to brain studies, IEEE Trans. Med. Imaging 33 (2014) 2332–2341.
[18] K. Gong, J. Yang, K. Kim, G. El Fakhri, Y. Seo, Q. Li, Attenuation correction for brain PET imaging using deep neural network based on Dixon and ZTE MR images, Phys. Med. Biol. (2018).
[19] S. Roy, W.-T. Wang, A. Carass, J.L. Prince, J.A. Butman, D.L. Pham, PET attenuation correction using synthetic CT from ultrashort echo-time MR imaging, J. Nucl. Med. 55 (2014) 2071–2077.
[20] A.P. Leynes, J. Yang, F. Wiesinger, S.S. Kaushik, D.D. Shanbhag, Y. Seo, T.A. Hope, P.E. Larson, 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, J. Nucl. Med. 59 (2018) 852–858.
[21] C. Bowles, C. Qin, C. Ledig, R. Guerrero, R. Gunn, A. Hammers, E. Sakka, D.A. Dickie, M.V. Hernández, N. Royle, et al., Pseudo-healthy image synthesis for white matter lesion segmentation, in: International Workshop on Simulation and Synthesis in Medical Imaging, Springer, 2016, pp. 87–96.
[22] A. Chartsias, T. Joyce, M.V. Giuffrida, S.A. Tsaftaris, Multimodal MR synthesis via modality-invariant latent representation, IEEE Trans. Med. Imaging (2017).
[23] J.P. Cohen, M. Luck, S. Honari, Distribution matching losses can hallucinate features in medical image translation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2018, pp. 529–536.

[24] A. Johansson, A. Garpebring, T. Asklund, T. Nyholm, CT substitutes derived from MR images reconstructed with parallel imaging, Med. Phys. 41 (2014).

[25] T. Joyce, A. Chartsias, S.A. Tsaftaris, Robust multi-modal mr image synthesis, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2017, pp. 347–355.

[26] D. Nie, R. Trullo, J. Lian, C. Petitjean, S. Ruan, Q. Wang, D. Shen, Medical image synthesis with context-aware generative adversarial networks, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2017, pp. 417–425.

[27] H. Van Nguyen, K. Zhou, R. Vemulapalli, Cross-domain synthesis of medical images using efficient location-sensitive deep network, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 677–684.

[28] J.M. Wolterink, A.M. Dinkla, M.H. Savenije, P.R. Seevinck, C.A. van den Berg, I. Išgum, Deep MR to CT synthesis using unpaired data, in: International Workshop on Simulation and Synthesis in Medical Imaging, Springer, 2017, pp. 14–23.

[29] D. Nie, R. Trullo, J. Lian, L. Wang, C. Petitjean, S. Ruan, Q. Wang, D. Shen, Medical image synthesis with deep convolutional adversarial networks, IEEE Trans. Biomed. Eng. 65 (2018) 2720–2730.

[30] Y. Huo, Z. Xu, S. Bao, A. Assad, R.G. Abramson, B.A. Landman, Adversarial synthesis learning enables segmentation without target modality ground truth, in: Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on, IEEE, 2018, pp. 1217–1220.

[31] A. Chartsias, T. Joyce, R. Dharmakumar, S.A. Tsaftaris, Adversarial image synthesis for unpaired multi-modal cardiac data, in: International Workshop on Simulation and Synthesis in Medical Imaging, Springer, 2017, pp. 3–13.

[32] Y. Hiasa, Y. Otake, M. Takao, T. Matsuoka, K. Takashima, A. Carass, J.L. Prince, N. Sugano, Y. Sato, Cross-modality image synthesis from unpaired data using cyclegan, in: International Workshop on Simulation and Synthesis in Medical Imaging, Springer, 2018, pp. 31–41.

[33] P. Costa, A. Galdran, M.I. Meyer, M. Niemeijer, M. Abràmoff, A.M. Mendonça, A. Campilho, End-to-end adversarial retinal image synthesis, IEEE Trans. Med. Imaging 37 (3) (2017) 781–791.

[34] R. Vemulapalli, H. Van Nguyen, S. Kevin Zhou, Unsupervised cross-modal synthesis of subject-specific scans, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 630–638.

[35] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: Advances in Neural Information Processing Systems, 2014, pp. 2672–2680.

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

[37] J.-Y. Zhu, T. Park, P. Isola, A.A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2223–2232.

[38] J.E. Iglesias, M. Modat, L. Peter, A. Stevens, R. Annunziata, T. Vercauteren, E. Lein, B. Fischl, S. Ourselin, A.D.N. Initiative, et al., Joint registration and synthesis using a probabilistic model for alignment of MRI and histological sections, Med. Image Anal. 50 (2018) 127–144.

[39] A. Jog, A. Carass, S. Roy, D.L. Pham, J.L. Prince, Random forest regression for magnetic resonance image synthesis, Med. Image Anal. 35 (2017) 475–488.

[40] F.J. Martinez-Murcia, J.M. Górriz, J. Ramírez, I.A. Illán, F. Segovia, D. Castillo-Barnes, D. Salas-Gonzalez, Functional brain imaging synthesis based on image decomposition and kernel modeling: Application to neurodegenerative diseases, Front. Neuroinform. 11 (2017) 65.

[41] D.H. Ye, D. Zikic, B. Glocker, A. Criminisi, E. Konukoglu, Modality propagation: coherent synthesis of subject-specific scans with data-driven regularization, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2013, pp. 606–613.

[42] A. Jog, S. Roy, A. Carass, J.L. Prince, Magnetic resonance image synthesis through patch regression, in: Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on, IEEE, 2013, pp. 350–353.

[43] A. Jog, A. Carass, S. Roy, D.L. Pham, J.L. Prince, MR image synthesis by contrast learning on neighborhood ensembles, Med. Image Anal. 24 (1) (2015) 63–76.

[44] N. Bayramoglu, M. Kaakinen, L. Eklund, J. Heikkila, Towards virtual h&e staining of hyperspectral lung histology images using conditional generative adversarial networks, in: Proceedings of the IEEE International Conference on Computer Vision Workshops, 2017, pp. 64–71.

[45] Y. Huang, L. Shao, A.F. Frangi, Cross-modality image synthesis via weakly coupled and geometry co-regularized joint dictionary learning, IEEE Trans. Med. Imaging 37 (3) (2018) 815–827.

[46] Y. Zhu, Y. Tang, Y. Tang, D.C. Elton, S. Lee, P.J. Pickhardt, R.M. Summers, Cross-domain medical image translation by shared latent Gaussian mixture model, 2020, arXiv preprint arXiv:2007.07230.

[47] M.-Y. Liu, T. Breuel, J. Kautz, Unsupervised image-to-image translation networks, in: I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), Advances in Neural Information Processing Systems 30, Curran Associates, Inc., 2019, pp. 700–708.

[48] H. Yang, J. Sun, A. Carass, C. Zhao, J. Lee, Z. Xu, J. Prince, Unpaired brain MR-to-CT synthesis using a structure-constrained cyclegan, in: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Springer, 2018, pp. 174–182.

[49] M.P. Heinrich, M. Jenkinson, M. Bhushan, T. Matin, F.V. Gleeson, M. Brady, J.A. Schnabel, MIND: Modality independent neighbourhood descriptor for multi-modal deformable registration, Med. Image Anal. 16 (7) (2012) 1423–1435.

[50] C. Wang, G. Papanastasiou, S. Tsaftaris, G. Yang, C. Gray, D. Newby, G. Macnaught, T. MacGillivray, Tpsdicyc: Improved deformation invariant cross-domain medical image synthesis, in: International Workshop on Machine Learning for Medical Image Reconstruction, Springer, 2019, pp. 245–254.

[51] C. Qin, B. Shi, R. Liao, T. Mansi, D. Rueckert, A. Kamen, Unsupervised deformable registration for multi-modal images via disentangled representations, in: International Conference on Information Processing in Medical Imaging, Springer, 2019, pp. 249–261.

[52] A. Chartsias, T. Joyce, G. Papanastasiou, S. Semple, M. Williams, D.E. Newby, R. Dharmakumar, S.A. Tsaftaris, Disentangled representation learning in cardiac image analysis, Med. Image Anal. 58 (2019) 101535.

[53] J. Kent, K. Mardia, The link between kriging and thin-plate splines, 1994.

[54] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, Y. Wei, Deformable convolutional networks, 2017, p. 3, CoRR, abs/1703.06211 1.

[55] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, P. Abbeel, Infogan: Interpretable representation learning by information maximizing generative adversarial nets, in: Advances in Neural Information Processing Systems, 2016, pp. 2172–2180.

[56] N.X. Vinh, J. Epps, J. Bailey, Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance, J. Mach. Learn. Res. 11 (Oct) (2010) 2837–2854.

[57] D. Newby, R. Forsythe, O. McBride, J. Robson, A. Vesey, R. Chalmers, P. Burns, O.J. Garden, S. Semple, et al., Aortic wall inflammation predicts abdominal aortic aneurysm expansion, rupture, and need for surgical repair, Circulation 136 (9) (2017) 787–797.

[58] A. Hore, D. Ziou, Image quality metrics: PSNR vs. SSIM, in: Pattern Recognition (ICPR), 2010 20th International Conference on, IEEE, 2010, pp. 2366–2369.

[59] G. Papanastasiou, V. González-Castro, C. Gray, R. Forsythe, Y. Sourgia-Koutraki, N. Mitchard, D.E. Newby, S. Semple, Multidimensional assessments of abdominal aortic aneurysms by magnetic resonance against ultrasound diameter measurements, in: Annual Conference on Medical Image Understanding and Analysis, Springer, 2017, pp. 133–143.

[60] P. Isola, J.-Y. Zhu, T. Zhou, A.A. Efros, Image-to-image translation with conditional adversarial networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1125–1134.

[61] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, in: International Conference on Learning Representations (ICLR), 2015.