Unsupervised Learning for Cross-Domain Medical Image Synthesis Using Deformation Invariant Cycle Consistency Networks

Chengjia Wang[1,2](✉), Gillian Macnaught[1,2], Giorgos Papanastasiou[2],
Tom MacGillivray[2], and David Newby[1,2]

[1] BHF Centre for Cadiovascular Science, University of Edinburgh, Edinburgh, UK
chengjia.wang@ed.ac.uk
[2] Edinburgh Imaging Facility QMRI, University of Edinburgh, Edinburgh, UK

Abstract. Recently, the cycle-consistent generative adversarial networks (CycleGAN) has been widely used for synthesis of multi-domain medical images. The domain-specific nonlinear deformations captured by CycleGAN make the synthesized images difficult to be used for some applications, for example, generating pseudo-CT for PET-MR attenuation correction. This paper presents a deformation-invariant CycleGAN (DicycleGAN) method using deformable convolutional layers and new cycle-consistency losses. Its robustness dealing with data that suffer from domain-specific nonlinear deformations has been evaluated through comparison experiments performed on a multi-sequence brain MR dataset and a multi-modality abdominal dataset. Our method has displayed its ability to generate synthesized data that is aligned with the source while maintaining a proper quality of signal compared to CycleGAN-generated data. The proposed model also obtained comparable performance with CycleGAN when data from the source and target domains are alignable through simple affine transformations.

Keywords: Synthesis · Deep learning · GAN · Unsupervised learning

1 Introduction

Modern clinical practices make cross-domain medical image synthesis a technology gaining in popularity. (In this paper, we use the term "domain" to uniformly address different imaging modalities and parametric configurations.) Image synthesis allows one to handle and impute data of missing domains in standard statistical analysis [1], or to improve intermediate step of analysis such as registration [2], segmentation [3] and disease classification [4]. Our application is to

C. Wang—This work was supported by British Heart Foundation.

A. Gooya et al. (Eds.): SASHIMI 2018, LNCS 11037, pp. 52–60, 2018.
https://doi.org/10.1007/978-3-030-00536-8_6

generate pseudo-CT images from multi-sequence MR data [5]. The synthesized pseudo-CT images can be further used for the purpose of PET-MR attenuation correction [6].

State-of-the-art methods often train a deep convolutional neural network (CNN) as image generator following the learning procedure of the generative adversarial network (GAN) [7]. Many of these methods require to use aligned, or paired, datasets which is hard to obtain in practice when the data can not be aligned through an affine transformation. To deal with unpaired cross-domain data, a recent trend is to leverage CycleGAN losses [8] into the learning process to capture high-level information translatable between domains. Previous studies have shown that CycleGAN is robust to unpaired data [9]. However, not all information encoded in CycleGAN image generators should be used due to very distinct imaging qualities and characteristics in different domains, especially different modalities. Figure 1 displays a representative example of CycleGAN based cross-modality synthesis where the real CT and MR data were acquired from the same patient. It can be seen that the shape and relative positions of the scanner beds are very different. This problem can be addressed as "domain-specific deformation". Because the generator networks can not treat the spatial deformation and image contents separately, CycleGAN encodes this information and reproduce it in the forward pass, which causes misalignment between the source and synthesized images. For some applications, such as generating pseudo-CT for attenuation correction of PET-MR data, this domain-specific deformation should be discarded. In the mean time, the networks should keep efficient information about appearences of the same anatomy in distinct domains. A popular strategy to solve this problem is performing supervised or semi-supervised learning with an extra mission, for example, segmentation [10], but this requires collection of extra ground truth.

In this paper, we present a deformation invariant CycleGAN (DicycleGAN) framework for cross-domain medical image synthesis. The architecture of the networks is inspired by the design of deformable convolutional network (DCN) [11]. We handle the different nonlinear deformations in different domains by integrating a modified DCN structure into the image generators and propose to use normalized mutual information (NMI) in the CycleGAN loss. We evaluate the proposed network using both multi-modality abdominal aortic data and multi-sequence brain MR data. The experimental results demonstrate the ability of our method to handle highly disparate imaging domains and generate synthesized images aligned with the source data. In the mean time, the quality of the synthesized images are as good as those generated by the CycleGAN model. The main contributions of this paper include a new DicycleGAN architecture which learns deformation-invariant correspondences between domains and a new NMI-based cycleGAN loss.

Table 1. Synthesis results of IXI dataset using undeformed T2 images.

Experiment	Model	MSE	PSNR	SSIM
PD to T2	Cycle	0.186 (0.08)	**27.35 (1.69)**	0.854 (0.03)
	Dicycle	**0.183 (0.09)**	26.49 (1.62)	**0.871 (0.03)**
T2 to PD	Cycle	**0.134 (0.02)**	**29.68 (1.61)**	**0.892 (0.03)**
	Dicycle	0.146 (0.03)	28.85 (1.59)	0.883 (0.02)

2 Method

A GAN framework using a image generator G to synthesize images of a target domain using images from a source domain, and a discriminator D to distinguish real and synthesized images. Parameters of G are optimized to confuse D, while D is trained at the same time for better binary classification performance to classify real and synthesized data. We assumes that we have n^A images $x^A \in \mathcal{X}^A$ from domain $\mathcal{X}^A$, and n^B images $x^B \in \mathcal{X}^B$. To generate synthesized images of domain $\mathcal{X}^B$ using images from $\mathcal{X}^A$, G and D are trained in the min-max game of the GAN loss $\mathcal{L}_{GAN}\left(G, D, \mathcal{X}^A, \mathcal{X}^B\right)$ [7]. When dealing with unpaired data, the original CycleGAN framework consists of two symmetric sets of generators $G^{A\to B}$ and $G^{B\to A}$ 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. The *cycle consistency* loss $\mathcal{L}_{cyc}\left(G^{A\to B}, D^A, G^{B\to A}, D^B, \mathcal{X}^A, \mathcal{X}^B\right)$, or $\mathcal{L}_{cyc}^{A,B}$, is used to keep the cycle-consistency between the two sets of networks. The loss of the whole CycleGAN framework $\mathcal{L}_{CycleGAN} = \mathcal{L}_{GAN}^{A\to B} + \mathcal{L}_{GAN}^{B\to A} + \lambda_{cyc}\mathcal{L}_{cyc}^{A,B}$. (In this paper we use the short expression $\mathcal{L}_{GAN}^{\mathcal{A}\to\mathcal{B}}$ to denote GAN loss $\mathcal{L}_{GAN}(G^{A\to B}, D^B, \mathcal{X}^A, \mathcal{X}^B)$). The image generator in the CycleGAN contains an input convolutional block, two down-sampling convolutional layers, followed by a few resnet blocks or a Unet structure, and two up-sampling transpose convolutional blocks before the last two convolutional blocks.

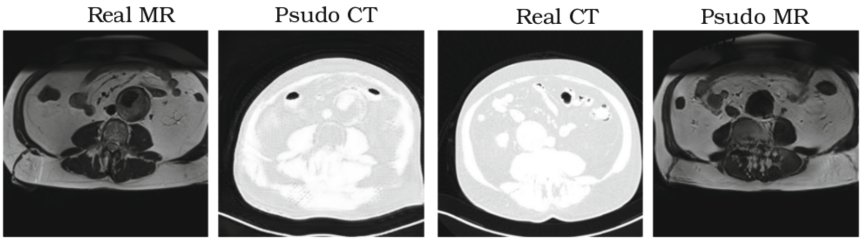

Fig. 1. Example of MR-CT synthesis using vanila CycleGAN.

DicycleGAN Architecture. In order to capture deformation-invariant information between domains, we introduce a modified DCN architecture into the image generators of CycleGAN, as shown in Fig. 2. The deformable convolution can be viewed as an atrous convolution kernel with trainable dilation rates and reinterpolated input feature map [11]. The spatial offsets of each point in the feature map is learned through a conventional convolution operation, followed by another convolution layer. This leads to a "Lasagne" structure consist of interleaved "offset convolution" and conventional convolution operations. We adopt this structure to the generators by inserting an offset convolutional operation (displayed in cyan in Fig. 2) in front of the input convolutional block, downsample convolutional blocks and the first resnet blocks. Note that this "offset" convolution only affects the interpolation of the input feature map rather than providing a real convolution result. Let θ_T denote the learnable parameters in the"offset" convolutional layers, and θ the rest parameters in image generator G. When training G, each input image x generates two output images: deformed output image $G_T(x) = G(x|\theta, \theta_T)$ and undeformed image $G(x) = G(x|\theta)$. The red and blue arrows in Fig. 2 indicate the computation flows for generating $G_T(x)$ and $G(x)$ in the forward passes. $G_T(x)$ is then taken by the corresponding discriminator D for calculation of GAN losses, and $G(x)$ is expected to be aligned with x.

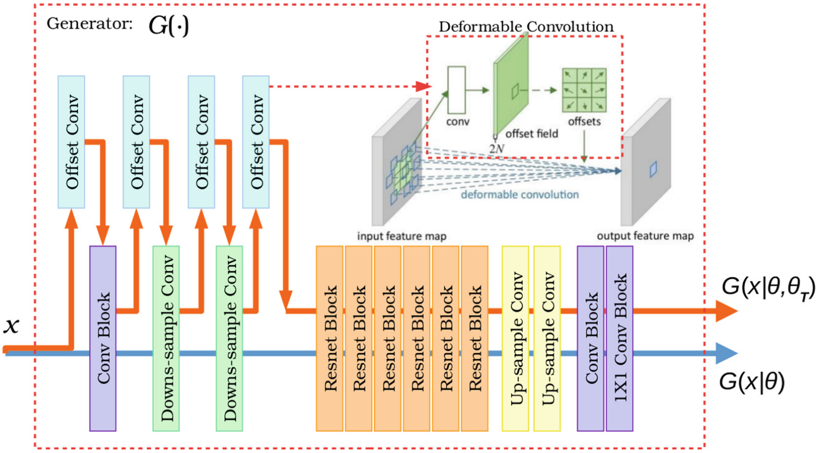

Fig. 2. Architecture of the proposed image generator $G(\cdot)$. Each input image x generates a deformed output $G(x|\theta, \theta_T)$ and an undeformed output $G(x|\theta)$ through two forward passes shown in red and blue. Demonstration of deformable convolution is obtained from [11]. (best viewed in color)

DicycleGAN Loss. DicycleGAN loss contains traditional GAN loss following the implementation of CycleGAN [8], but also includes an image alignment loss and a new cycle consistency loss. For the GAN loss $L_{GAN}^{A\rightarrow B}$, the image generator $G^{A\rightarrow B}$ is trained to minimize $\left(D^B\left(G_T^{A\rightarrow B}\left(x^A\right)\right)-1\right)^2$ and D^B is trained to minimize $\left(D^B(x^B)-1\right)^2 + D^B\left(G_T^{A\to B}(x^A)\right)^2$. The same formulation is used to calculate $\mathcal{L}_{GAN}^{B\to A}$ defined on $G^{B\to A}$ and D^A. Note that the GAN loss is calculated based on the deformed synthesized images. As the undeformed outputs of generators are expected to be aligned with the input images, we propose to use a information loss based on normalized mutual information (NMI). NMI is a popular metric used for image registration. It varies between 0 and 1 indicating alignment of two clustered images [12]. The image alignment loss is defined as:

$$\mathcal{L}_{align}^{A,B} = 2 - NMI\left(x^A, G^{A\to B}\left(x^A\right)\right) - NMI\left(x^B, G^{B\to A}\left(x^B\right)\right). \quad (1)$$

Based on the proposed design of image generators, the cycle two types of cycle consistency losses. The undeformed cycle consistency loss is defined as:

$$\mathcal{L}_{cyc}^{A,B} = \|G^{B\to A}\left(G^{A\to B}\left(x^A\right)\right) - x^A\|_1 + \|G^{A\to B}\left(G^{B\to A}\left(x^B\right)\right) - x^B\|_1. \quad (2)$$

Beside $\mathcal{L}_{cyc}$, the deformation applied to the synthesized image should be also cycle consistent. Here we defined a deformation-invariant cycle consistency loss:

$$\mathcal{L}_{dicyc}^{A,B} = \|G_T^{B\to A}\left(G_T^{A\to B}\left(x^A\right)\right) - x^A\|_1 + \|G_T^{A\to B}\left(G_T^{B\to A}\left(x^B\right)\right) - x^B\|_1. \quad (3)$$

To perform image synthesis between domains $\mathcal{X}^A$ and $\mathcal{X}^B$, we use the deformed output images $G_T^{A\to B}$ and $G_T^{B\to A}$ to calculate the GAN loss. The full loss of DicycleGAN is:

$$\mathcal{L}_{DicycleGAN} = \mathcal{L}_{GAN}^{A\to B} + \mathcal{L}_{GAN}^{B\to A} + \lambda_{align}\mathcal{L}_{align}^{A,B} + \lambda_{cyc}\mathcal{L}_{cyc}^{A,B} + \lambda_{dicyc}\mathcal{L}_{dicyc}^{A,B}. \quad (4)$$

Figure 3 provides a demonstration of computing the all the losses discussed above using outputs of the corresponding DicycleGAN generators and discriminators.

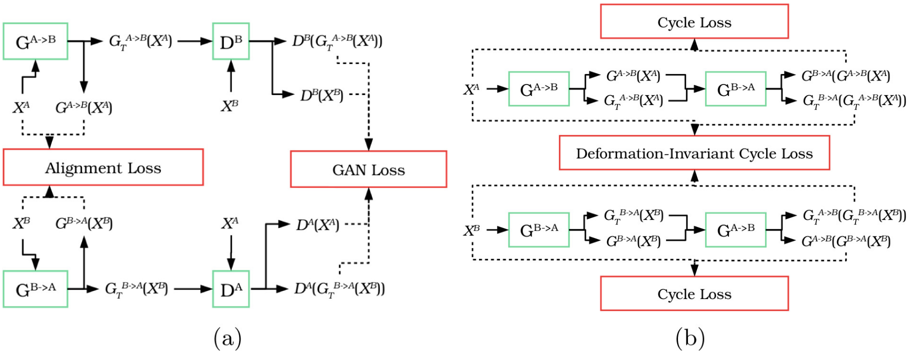

Fig. 3. Calculation of losses in DicycleGAN. (a) shows GAN and image alignment losses: the undeformed output of the image generators are used for alignment losses, and the deformed outputs for GAN losses. (b) shows the Cycle consistency losses.

3 Experiments

Evaluation Metrics. The most widely used quantitative evaluation metrics for cross-domain image synthesis are: mean squared error (MSE), peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM). Given a volume x^A and a target volume x^B, the MSE is computed as: $\frac{1}{N}\sum_1^N \left(x^B - G^{A\rightarrow B}(x^A)\right)^2$, where N is number of voxels in the volume. PSNR is calculated as: $10\log_{10}\frac{\max_B^2}{MSE}$. 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 [13].

Datasets. We use the Information eXtraction from Images (IXI) dataset[1] which provides co-registered multi-sequence skull-stripped MR images collected from multiple sites. Due to the limited storage space, here we selected 66 proton density (PD-) and T2-weighted volumes, each volume contains 116 to 130 2D slices. We use 38 pairs for training and 28 pairs for evaluation of synthesis results. Our image generators take 2D axial-plane slices of the volumes as inputs. During the training phase, we resample all volumes to a resolution of $1.8 \times 1.8 \times 1.8\,\mathrm{mm}^3/voxel$, then crop the volumes to 128×128 pixel images. As the generators in both CycleGAN and DicycleGAN are fully convolutional, the predictions are performed on full size images. All the images are normalized with their mean and standard deviation. We also used a private dataset contains 40 multi-modality abdominal T2*-weighted images and CT images collected from 20 patients with abdominal aortic aneurysm (AAA) in a clinical trial. All images are resampled to a resolution of $1.56 \times 1.56 \times 5\,\mathrm{mm}^3/voxel$, and the axial-plane slices trimmed to 192×192 pixels. It is difficult to non-rigidly register whole abdominal images to calculate the evaluation metrics, but the aorta can be rigidly aligned to assess the performance of image synthesis. The anatomy of the aorta have previously been co-registered and segmented by 4 clinical researchers.

Implementation Details. We used image generators with 9 Resnet blocks. All parameters of, or inherit from, vanilla CycleGAN are taken from the original paper. For the DicycleGAN, we set $\lambda_{cyc} = \lambda_{dicyc} = 10$ and $\lambda_{align} = 0.9$. The models were trained with Adam optimizer [14] 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 $\mathcal{L}_{DicycleGAN}$ stops decreasing for 10 epochs, the training will move to the learning rate decaying stage; this tolerance is set to 20 epochs in the second 100 epochs.

Experiments Setup. In order to quantitatively evaluate robustness of our model to the domain-specific local distortion, we applied an arbitrary non-linear distortion to the T2-weighted images of IXI. Synthesis experiments were performed between the PD-weighted data and undeformed, as well as the deformed

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

T2-weighted data. When using deformed T2-weighted images, the ground truth were generated by applying the same nonlinear deformation to the source PD-weighted images. We trained the CycleGAN and DicycleGAN using unpaired, randomly selected slices. The training images were augmented using aggressive flips, rotations, shearing and translations so that CycleGAN can be robust. In the test stage, the three metrics introduced above were computed. For our private dataset, the metrics were computed within the segmented aortic anatomy excluding any other imaged objects because all the three metrics require to be calculated on aligned images.

4 Results

Tables 1 and 2 present the PD-T2 co-synthesis results using undeformed and deformed T2-weighted images. In addition, Fig. 4 provides an example showing the synthesized images generated by CycleGAN and DicycleGAN. CycleGAN encoded the simulated domain-specific deformation, whether applied to source or target domain, and combined this deformation into the synthesized images. This leads to misalignment of source and synthesized images. The quantitative results show that our DicycleGAN model produced comparable results with CycleGAN when there is no domain-specific distortions, but it achieved remarkable performance gain when the source and target images can not be aligned through an affine transformation. This is because of the deformed synthesized images generated by CycleGAN which lead to severe misalignments between the source and synthesized images.

The cross-modality synthesis results are shown in Table 3. The discrepancy between the two imaging modalities can be shown by the different relative positions between the imaged objects and the beds. CycleGAN encoded this information in the image generators as shown earlier in Fig. 1.

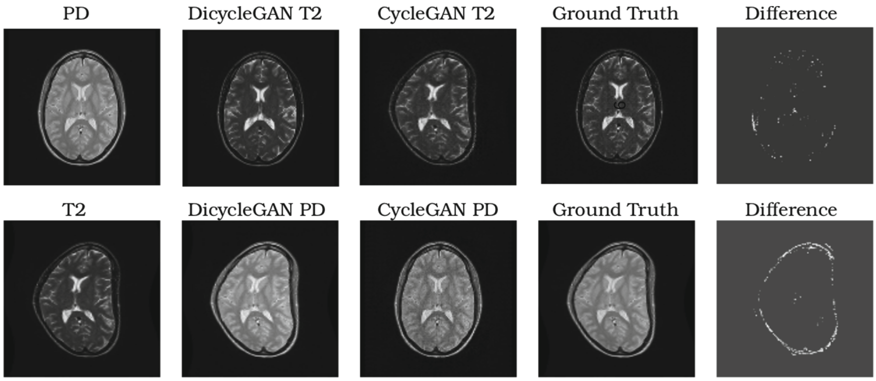

Fig. 4. Example of synthesized images generated by CycleGAN and DicycleGAN, compared to the ground truths. The ground truth of the deformed source image is generated by applying the same arbitrary deformation to the original target image.

Table 2. Synthesis results of IXI dataset using deformed T2 images.

Experiment	Model	MSE	PSNR	SSIM
PD to T2	Cycle	0.586 (0.25)	19.52 (1.62)	0.6081 (0.12)
	Dicycle	**0.145 (0.02)**	**22.32 (1.29)**	**0.7842 (0.03)**
T2 to PD	Cycle	0.561 (0.22)	19.42 (1.61)	0.6001 (0.11)
	Dicycle	**0.141 (0.02)**	**22.86 (1.31)**	**0.7714 (0.02)**

Table 3. Multi-modality synthesis results using private dataset.

Experiment	Model	MSE	PSNR	SSIM
T2* to CT	Cycle	0.516 (0.19)	18.32 (1.82)	0.5716 (0.15)
	Dicycle	**0.287 (0.11)**	**23.71 (1.17)**	**0.7122 (0.03)**
CT to T2*	Cycle	0.521 (0.22)	19.12 (1.60)	0.5818 (0.12)
	Dicycle	**0.299 (0.08)**	**22.66 (1.11)**	**0.7556 (0.02)**

5 Conclusion and Discussion

We propose a new method for cross-domain medical image synthesis, called DicycleGAN. Compared to the vanilla CycleGAN method, we integrate DCN layers into the image generators and reinforce the training process with deformation-invariant cycle consistency loss and NMI-based alignment loss. Results obtained from both multi-sequence MR dataset and our private multi-modality abdominal dataset shows that our model achieved comparable performance with CycleGAN when the source and target data can be aligned with an affine transformation. Our model achieved obvious performance gain compared to CycleGAN when there are domain-specific nonlinear deformations. A possible future application of DicycleGAN is multi-modal image registration.

References

1. van Tulder, G., de Bruijne, M.: Why does synthesized data improve multi-sequence classification? In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9349, pp. 531–538. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24553-9_65
2. Iglesias, J.E., Konukoglu, E., Zikic, D., Glocker, B., Van Leemput, K., Fischl, B.: Is synthesizing MRI contrast useful for inter-modality analysis? In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013. LNCS, vol. 8149, pp. 631–638. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-40811-3_79
3. Roy, S., Carass, A., Prince, J.: A compressed sensing approach for MR tissue contrast synthesis. In: Székely, G., Hahn, H.K. (eds.) IPMI 2011. LNCS, vol. 6801, pp. 371–383. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-22092-0_31

4. Li, R., et al.: Deep learning based imaging data completion for improved brain disease diagnosis. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) MICCAI 2014. LNCS, vol. 8675, pp. 305–312. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10443-0_39
5. Nie, D., et al.: Medical image synthesis with context-aware generative adversarial networks. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10435, pp. 417–425. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66179-7_48
6. Wagenknecht, G., Kaiser, H.J., Mottaghy, F.M., Herzog, H.: MRI for attenuation correction in PET: methods and challenges. Magn. Resonance Mater. Phys. Biol. Med. **26**(1), 99–113 (2013)
7. Goodfellow, I., et al.: Generative adversarial nets. In: Advances in Neural Information Processing Systems, pp. 2672–2680 (2014)
8. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. arXiv preprint arXiv:1703.10593 (2017)
9. Wolterink, J.M., Dinkla, A.M., Savenije, M.H.F., Seevinck, P.R., van den Berg, C.A.T., Išgum, I.: Deep MR to CT synthesis using unpaired data. In: Tsaftaris, S.A., Gooya, A., Frangi, A.F., Prince, J.L. (eds.) SASHIMI 2017. LNCS, vol. 10557, pp. 14–23. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-68127-6_2
10. Huo, Y., Xu, Z., Bao, S., Assad, A., Abramson, R.G., Landman, B.A.: Adversarial synthesis learning enables segmentation without target modality ground truth. arXiv preprint arXiv:1712.07695 (2017)
11. Dai, J., et al.: Deformable convolutional networks. CoRR, abs/1703.06211 **1**(2), 3 (2017)
12. Vinh, N.X., Epps, J., Bailey, J.: Information theoretic measures for clusterings comparison: variants, properties, normalization and correction for chance. J. Mach. Learn. Res. **11**(Oct), 2837–2854 (2010)
13. Larkin, K.G.: Structural similarity index SSIMplified (2015)
14. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)