Adversarial Image Synthesis for Unpaired Multi-Modal Cardiac Data

Agisilaos Chartsias*[1], Thomas Joyce*[1], Rohan Dharmakumar[2], and Sotirios A. Tsaftaris[1]

[1]Institute for Digital Communications, School of Engineering, University of Edinburgh, West Mains Rd, Edinburgh EH9 3FB, UK, [2]Cedars Sinai Medical Center Los Angeles CA, USA
agis.chartsias@ed.ac.uk

Abstract. This paper demonstrates the potential for synthesis of medical images in one modality (e.g. MR) from images in another (e.g. CT) using a CycleGAN [25] architecture. The synthesis can be learned from unpaired images, and applied directly to expand the quantity of available training data for a given task. We demonstrate the application of this approach in synthesising cardiac MR images from CT images, using a dataset of MR and CT images coming from different patients. Since there can be no direct evaluation of the synthetic images, as no ground truth images exist, we demonstrate their utility by leveraging our synthetic data to achieve improved results in segmentation. Specifically, we show that training on both real and synthetic data increases accuracy by 15% compared to real data. Additionally, our synthetic data is of sufficient quality to be used alone to train a segmentation neural network, that achieves 95% of the accuracy of the same model trained on real data.

Keywords: Synthesis, MR, CT, Cardiac, Deep Learning, GAN

1 Introduction

Medical imaging research has benefited significantly from the application of modern deep learning techniques. Yet, often the very best deep learning results outwith medical imaging are achieved when large labelled datasets are available. This is difficult in the medical setting, as medical data is often very sparsely labelled (generally requiring labelling by experts), expensive to obtain, and has to respect patient anonymity constraints. All of these factors make large labelled datasets rare in medical image analysis, and thus investigation into methods for mitigating this restriction are valuable.

When attempting to develop a model for a new task, it is common for only a limited quantity of labelled data in the modality of interest to exist. However, the same anatomy may have been imaged in other individuals and in other modalities, and then carefully labelled by experts. The fact that this labelled

* These authors contributed equally.

data is not in the 'correct' modality means it is not immediately useful, but the ability to make use of these auxiliary labelled datasets would be extremely valuable, potentially enlarging the pool of labelled data many-fold.

In this paper we propose a pipeline for directly transforming auxiliary labelled data into the modality of interest (the "target modality"). We demonstrate that a small set of labelled data in the target modality can be used as a bootstrap, allowing us to convert labelled data from other modalities into the desired modality and expand the dataset. Additionally, this synthetic data consists of new examples not derived from existing examples, and potentially containing beneficial new anatomical and topological information from the auxiliary data. We show that this larger and more diverse dataset can then be used to train an improved model for the task at hand. Here, we demonstrate this for myocardial segmentation. However, as the data only needs to be of the same anatomy (not necessarily from the same individuals for example), the method can potentially expand the available training data for many tasks. Moreover, the method is especially suitable for cardiac use, as it does not require co-registered data.

The pipeline for our approach is as follows: firstly, we perform a view alignment step, transforming the auxiliary data so that the scale, position and viewing angle is broadly the same as in the target modality (Section 3.1). Secondly, we make use of a CycleGAN [25] architecture for unpaired image synthesis. This uses adversarial training to overcome the need for aligned pairs of images in the source and target modalities, and learns to transform data from one modality to the other. Once trained, we use the learned transformation to convert all the auxiliary data into synthetic data in the target modality (Section 3.2). A schematic overview of our approach is given in Figure 1.

To evaluate this approach we apply our method to cardiac synthesis, generating cardiac MR images from cardiac CT images (Section 4). Directly quantitatively assessing the quality of synthetic data when no ground truth exists is very challenging. We demonstrate the synthetic data's utility by showing it significantly improves results in a segmentation task.

Specifically, this paper makes the following contributions: 1. Introduction of a flexible pipeline for transforming labelled data in auxiliary modalities into labelled data in the modality of interest. 2. A demonstration that augmenting real data with this synthetic data significantly improves performance in a segmentation task. 3. Comparison of our synthetic augmentation with standard augmentation, showing the synthesis approach to be favourable. 4. Demonstration of a recommended approach, which combines both synthesis and augmentation, and results in the best performance overall.

2 Previous Work

There has been very little previous work on learning-based methods for cardiac synthesis. Existing approaches have focused on combining electro-mechanical models of the heart's motion with template real images for generating image simulations [1, 15] and have been recently extended for simulation of pathologic

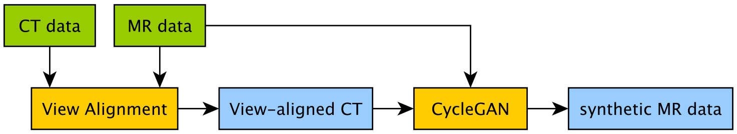

Fig. 1. A high-level schematic of the synthesis pipeline for the cardiac data. he CycleGAN also produces synthetic CT images but here we only use the synthetic MR.

from healthy cases [5] and for multimodal image simulation of both pathologic and healthy images in MR modalities [24].

Our work is based on learning a transformation function between images in order to transfer anatomical information from a source to a target modality. Similar methods have been proposed for cross-modal synthesis of brain images [4, 7, 10, 18, 22]. However, these are made possible by the availability of co-registered multimodal datasets, which allow a mapping from one modality to another to be directly learned using supervised techniques. In the cardiac domain, such registered multimodal datasets are harder to create. This is in part due to several unique challenges that cardiac data presents. Many of these difficulties result from the fact that the heart is an active moving muscle during the data acquisition session, and thus imaging it is more difficult than imaging the brain, or bones, which are essentially static relative to the body. In addition, the fact that the heart is moving makes it very difficult to produce co-registered images of the heart in different imaging modalities, and registration is often a complex non-linear post-processing step [19].

Cardiac synthesis methods have been explored for super-resolution (i.e. spatial up-sampling) in [12]. These methods can be learned by creating a low resolution version of a dataset, and then learning to synthesise the original resolution, again admitting a supervised approach. Recently, cardiac super-resolution has been enhanced by incorporating a shape prior in the learning process [13].

Furthermore, super-resolution has been coupled with cross-modal synthesis in a dictionary learning approach, with the addition of unpaired data in the learning process to improve the quality of results [8], proposing a weakly supervised learning approach. Unpaired data has also been used for cross-modal synthesis in an optimisation scheme [23], treating the problem as unsupervised learning. However, [23] focuses only on brain images, and does not address cardiac.

Unsupervised learning, for example learning image transformations with no ground-truth target images, has been revolutionalised by the introduction of adversarial training of neural networks [6, 16]. Adversarial learning was used for image style transformation in [25], and this method is directly applicable to cardiac data, where there is a lack of paired data.

Although synthesis offers a flexible approach that can be directly applied to expand available data, it is still important to weigh synthesis up, critically, against other approaches. Synthetic data has been used previously to improve segmentation [9] and classification algorithms [21], and, as there is no direct way to measure accuracy when ground truth images do not exist, the value of synthesis should be measured by considering how well it achieves these aims. However, this means that synthesis should also be compared with alternative methods for achieving these same goals.

In this paper we demonstrate the utility of synthesis for improving segmentation via enlarging the set of available training data. Besides synthesis, a dataset can also be expanded using simple geometric augmentation, for example by rotating and reflecting the images. Although simple transformation based augmentation is commonly used to improve results on cardiac segmentation [20, 14], this approach produces derivative examples, and does not benefit from the existence of auxiliary data, which could potentially provide additional real anatomical examples. In our experiments (Section 4) we directly compare this standard data augmentation with our synthesis approach, and, as the approaches are not mutually exclusive, also explore combining both approaches.

3 Method Details

We now give step-by-step details of our method, describing the view alignment, the training of the CycleGAN and the generation of the synthetic data.

3.1 View Alignment

In the view alignment step we make the CT and MR image sets broadly similar in terms of structure. Specifically, we aim to make the layout of the images (the position and size of the anatomy for example) not informative as to the dataset from which the image came. Preventing this is important in order to ensure the adversarial training is effective, otherwise the discriminator may learn to differentiate between real and synthetic data by attending to structural differences, rather than intensity statistics. However, the alignment only needs to be approximate, and any simple registration approach should suffice. Here we make use of the multiple labels on the data, using them to approximate the affine transformation that, for a given MR and CT volume, when applied to the CT volume, maximises alignment with the MR volume. After this crude alignment, any points in the new CT volume that correspond to points outside of the original CT volume are set to 0. Additionally, any points in the MR volume that correspond to points outside of the original CT volume are also set to 0. This again is performed to make the volumes structurally similar, to aid the adversarial training.

3.2 Transform Learning with CycleGAN

Since the images are not paired, learning to transform from MR to CT is not straightforward. However, a recent adversarial approach to this difficult task is the CycleGAN [25]: an adversarialy trained deep network which simultaneously learns transformations between two datasets containing the same information,

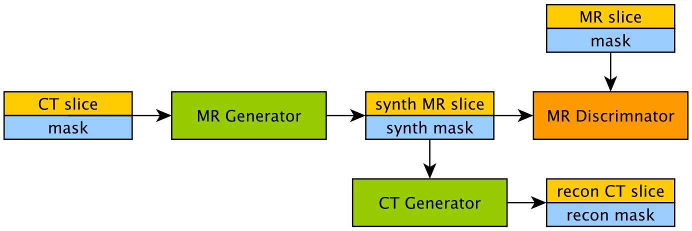

Fig. 2. Unfolded CycleGAN [25] training for CT to MR synthesis: a CT image with its segmentation mask is mapped to a synthetic MR and mask by a generator network $F : [CT, Mask] \rightarrow [MR, Mask]$. An MR discriminator then tries to discriminate real from synthetic MR. The CT and Mask are also reconstructed form the synthetic MR by a second generator network $G : [MR, Mask] \rightarrow [CT, Mask]$, which aims to reconstruct the original CT exactly. The generator learns both by trying to fool the discriminator, and by minimising the discrepancy between the real CT and its reconstruction.

but differently represented. It is powerful since it does not require paired training data, but instead learns via the use of both a discriminator and a cycle loss.

Specifically, a transform $F : A \rightarrow B$ is learned from dataset A to dataset B to produce synthetic B data y^B from real A data x^A, i.e. $y^B = F(x^A)$. Transform F aims to fool a discriminator D_B, which is simultaneously learning to discriminate between real and synthetic B data. Additionally, the synthetic B data is then transformed back into its original modality by a second learned transform $G : B \rightarrow A$ going in the other direction $x^A = G(y^B)$. This allows for an additional cost to be included in the training: data mapped from A to B, then back to A, should be as close as possible to the original A. That is $G(F(x^A))$ should be equal to x^A. This cycle loss gives the model its name.

In our case, the CycleGAN learns to transform a CT image into a synthetic MR image that cannot be recognised as synthetic by a discriminator network. At the same time, the synthetic MR image must be able to be accurately converted back into a CT image, as similar as possible to the original CT image, via another learned transformation. Thus, the synthetic MR image, whilst appearing realistic, must also retain relevant information from the CT. This encourages the synthetic MR to contain the same anatomy as is present in the input CT.

Adversarial training of deep neural networks is a challenging task, sensitive to many variables, and increasing the accessibility of the approach is an active area of research [2, 3]. Initially, we applied the CycleGAN directly to the MR and CT images. However, we found that although the resulting images were promising in terms of realism, the myocardium in the synthetic image was often not in the same place as in the input image. As a result, our synthetic MR data had no accurate labels, as we could not assume the label was the same as in the input image. To mitigate this issue we included both the mask of the myocardium and the image as two channel inputs to the CycleGAN, such that it learned to transform CT images and their corresponding myocardium segmentation mask into realistic MR images and corresponding segmentation masks. This did not stop the anatomy shifting during the transformation, but meant that we still had accurate (synthetic) labels for the synthetic images. A schematic of this approach can be seen in Figure 2.

3.3 Synthesis

We apply the mapping learned with the CycleGAN to the view-aligned CT data and the CT masks, producing a synthetic MR image and mask for every CT image in the dataset. The result is a synthetic labelled data set of MR cardiac images, which can be used for any task of interest.

4 Experiments

In this section we examine the effect of synthetic results in the accuracy of myocardium segmentation. We train a segmentation model, detailed in Section 4.1, on various combinations of synthetic and real data, with and without augmentation and report the dice coefficient on 3-fold cross validation.

4.1 Segmentation

To segment the images, we train a neural network with an architecture similar to the U-Net [17]. Specifically, the network consists of 3 downsample and 3 upsample blocks with skip connections between each block of equal size filters. This architecture was chosen as similar fully convolutional networks have been shown to achieve state of the art results in various segmentation tasks, including cardiac, and U-Net is a standard benchmark approach. Here we have not specifically optimised the architecture or hyperparameters for the segmentation task being considered, since the aim is to evaluate the synthetic results. Our model is implemented in Keras[1] and trained using Adam [11] with batch-size 16 and an early stopping criterion, based on the validation data, to avoid overfitting.

4.2 Data

We use 40 anonymised volumes, of which 20 are cardiac CT/CTA and 20 are cardiac MRI, kindly made available by the authors of [26, 27]. The CT/CTA data were acquired at Shanghai Shuguang Hospital, China, using routine cardiac CT angiography. The slices were acquired in the axial view. The inplane resolution is about $0.78 \times 0.78mm$ and the average slice thickness is $1.60mm$. The MRI data were acquired at St. Thomas hospital and Royal Brompton Hospital, London, UK, using 3D balanced steady state free precession (b-SSFP) sequences, with about $2mm$ acquisition resolution at each direction and reconstructed (resampled) into about $1mm$. The data contains static 3D images, acquired at different time points relative to the systole and diastole. All the data has manual segmentation of the seven whole heart substructures. However, in our segmentation experiments we only use the labels for the myocardium of the left ventricle.

[1] https://keras.io

4.3 Data Preprocessing

We centered the anatomy (the bounding box of the labeled anatomical regions) within the MR volumes, and trimmed each volume to 232×232, padding with 0s where necessary, but maintaining the native resolution. Then, for each volume, we clipped the top 1% of pixel values and re-scaled the values to $[-1, 1]$. Finally, we removed slices that did not contain myocardium, resulting in 20 volumes with an average of 41 slices per volume (816 slices in total). For the cardiac CT data no centering or trimming was necessary, as the data is aligned with the MR data in the view alignment step of Section 3.1. However, we again clipped the top 1% of values, and scaled the values to $[-1, 1]$.

4.4 Experiment Details

Below we detail the five experiments we used to evaluate the quality of the synthesised cardiac MR data. We repeated all experiments on three different splits of the data, each time training a CycleGAN on 15 MR and 15 CT volumes, and then training the segmentation network described in Section 4.1. In every split, the 5 MR volumes used for testing the segmentation network were excluded, as were the 5 CT volumes which were aligned with them in the view alignment step. Thus the final test volumes have not been used anywhere in the pipeline. Out of the remaining 15 MR volumes, we used 10 for training and 5 for validation.
Real: Firstly, as a baseline we train the segmentation network on 10 real MR volumes, using the other 5 MR volumes for validation, and obtain a mean dice coefficient of 0.613 on the test set.
Synthetic: Secondly, to directly evaluate the quality of the synthetic data, we train the segmentation network on 10 synthetic volumes, validating on 5 synthetic volumes. We then test the final model on the 5 real MR volumes and obtained a dice coefficient of 0.580.
Real and Synthetic: Next we combine the real and synthetic data and train our segmentation network on a total of 25 volumes (10 real and 15 synthetic), again using 5 real volumes for validation. This combined training gives a performance gain of $\sim 15\%$ compared to training on real data alone.
Augmented Real: Next we augment the real data using horizontal and vertical flips generating a total training set of 25 volumes (10 real 15 flipped) to allow direct comparison with synthetic augmentation.
Augmented Real and Synthetic: Finally, we combine the real and synthetic training data, and also use horizontal and vertical flips to expand the data to double the size. This results in 50 training volumes, and we again use 5 real volumes for validation during training.

4.5 Results

All results are presented side-by-side in Table 1. In addition, in Figure 3 we provide examples of our synthetic results. The first observation is that using just the synthetic data is almost as good as using the real data, in terms of resulting

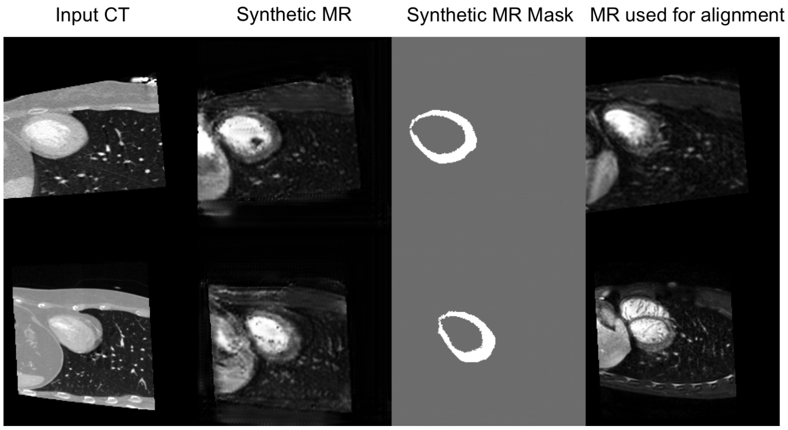

Fig. 3. Two examples of MR synthesis. From left to right it is shown, the real CT image, the resulting synthetic MR image, the synthetic segmentation mask and finally the real MR image of the volume to which the real CT volume was aligned in the view alignment step. Note that the shape and position of the myocardium is similar but not identical between the CT input and corresponding synthetic MR output. Also, observe that in the upper row the synthetic data contains a dark artifact within the ventricle.

segmentation, only resulting in a 5% loss of accuracy and this difference is not statistically significant at the 5% level. This is likely the result of small errors present in the synthetic images. Next, it is informative to compare real data with standard augmentations against the combined real and synthetic data. In both cases the segmentation algorithm was trained on 25 volumes, including the same 10 real volumes, and both approaches improve the final segmentation accuracy with synthetic and geometric augmentation leading to 14.8% and 10.3% improvements respectively. Finally, when the real and synthetic data is combined, and geometric augmentations are also applied, the greatest improvement is seen, with a 16.1% increase in accuracy over the baseline.

The difference in performance between the real and synthetic data, and just the real data is significant at the 5% level, as is the difference between the real and synthetic data and the augmented real data. Further, adding augmentation to the real and synthetic data does not lead to a statistically significant improvement.

5 Discussion and Conclusion

We have demonstrated that it is possible to produce synthetic cardiac data from unpaired images coming from different individuals. Moreover, we demonstrate that these synthetic images are accurate enough to be of significant benefit for further tasks, either used alone or to enlarge existing data sets. Specifically, we have shown that it is possible to produce synthetic cardiac MR images from cardiac CT images, and that these images can be used to improve the accuracy

Table 1. Dice scores for U-Nets trained on various data combinations. In all cases the model is evaluated on real MR images.

training data	split 1	split 2	split 3	average	relative to real
just synthetic	0.553	0.516	0.672	0.580	0.946
just real	0.584	0.613	0.642	0.613	1.000
augmented real	0.632	0.685	0.711	0.676	1.103
real and synthetic	**0.657**	0.699	**0.757**	0.704	1.148
augmented real and synthetic	0.650	**0.738**	0.748	**0.712**	**1.161**

of a segmentation algorithm by 16% when used in combination with standard geometric augmentation techniques. We also demonstrated that the synthetic data alone was sufficient to train a segmentation algorithm only 5% less accurate than the same algorithm trained entirely on real data.

As can be seen in the results, the largest gains are made when the synthetic data is included in the training set, suggesting that new anatomy, containing additional examples of real structure and natural local variations, being introduced from the auxiliary data is most beneficial for improving results.

Acknowledgements This work was supported in part by the US National Institutes of Health (2R01HL091989-05) and UK EPSRC (EP/P022928/1). We thank NVIDIA for donating a Titan X GPU.

1. Alessandrini, M., De Craene, M., Bernard, O., Giffard-Roisin, S., Allain, P., Waechter-Stehle, I., Weese, J., Saloux, E., Delingette, H., Sermesant, M.: A pipeline for the generation of realistic 3D synthetic echocardiographic sequences: methodology and open-access database. IEEE TMI 34(7), 1436–1451 (2015)
2. Arjovsky, M., Chintala, S., and Bottou, L.: Wasserstein gan. preprint arXiv: 1701.07875 (2017)
3. Berthelot, D., Schumm, T., and Metz, L.: BEGAN: Boundary Equilibrium Generative Adversarial Networks. preprint arXiv:1703.10717 (2017)
4. Cordier, N., Delingette, H., Lê, M., and Ayache, N.: Extended Modality Propagation: Image Synthesis of Pathological Cases. IEEE TMI 35(12), 2598–2608 (2016)
5. Duchateau, N., Sermesant, M., Delingette, H., and Ayache, N.: Model-based generation of large databases of cardiac images: synthesis of pathological cine MR sequences from real healthy cases. IEEE TMI (2017)
6. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y.: Generative adversarial nets. In: NIPS, pp. 2672–2680 (2014)
7. Huang, Y., Beltrachini, L., Shao, L., and Frangi, A.F.: Geometry Regularized Joint Dictionary Learning for Cross-Modality Image Synthesis in Magnetic Resonance Imaging. In: SASHIMI, pp. 118–126 (2016)
8. Huang, Y., Shao, L., and Frangi, A.F.: Simultaneous Super-Resolution and Cross-Modality Synthesis of 3D Medical Images using Weakly-Supervised Joint Convolutional Sparse Coding. preprint arXiv:1705.02596 (2017)

9. Iglesias, J.E., Konukoglu, E., Zikic, D., Glocker, B., Van Leemput, K., and Fischl, B.: Is synthesizing MRI contrast useful for inter-modality analysis? In: MICCAI, pp. 631–638 (2013)
10. Jog, A., Carass, A., Roy, S., Pham, D.L., and Prince, J.L.: Random forest regression for magnetic resonance image synthesis. Medical Image Analysis 35, 475–488 (2017)
11. Kingma, D., and Ba, J.: Adam: A method for stochastic optimization. preprint arXiv: 1412.6980 (2014)
12. Oktay, O.: Multi-input Cardiac Image Super-Resolution Using Convolutional Neural Networks. In: MICCAI, pp. 246–254 (2016)
13. Oktay, O., Ferrante, E., Kamnitsas, K., Heinrich, M., Bai, W., Caballero, J., Guerrero, R., Cook, S., Marvao, A. de, O'Regan, D.: Anatomically Constrained Neural Networks (ACNN): Application to Cardiac Image Enhancement and Segmentation. preprint arXiv:1705.08302 (2017)
14. Poudel, R.P., Lamata, P., and Montana, G.: Recurrent Fully Convolutional Neural Networks for Multi-slice MRI Cardiac Segmentation. preprint arXiv:1608.03974 (2016)
15. Prakosa, A., Sermesant, M., Delingette, H., Marchesseau, S., Saloux, E., Allain, P., Villain, N., and Ayache, N.: Generation of synthetic but visually realistic time series of cardiac images combining a biophysical model and clinical images. IEEE TMI 32(1), 99–109 (2013)
16. Radford, A., Metz, L., and Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. preprint arXiv:1511.06434 (2015)
17. Ronneberger, O., Fischer, P., and Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. MICCAI (2015)
18. Sevetlidis, V., Giuffrida, M.V., and Tsaftaris, S.A.: Whole Image Synthesis Using a Deep Encoder-Decoder Network. In: SASHIMI, pp. 97–107 (2016)
19. Tavakoli, V., and Amini, A.A.: A survey of shaped-based registration and segmentation techniques for cardiac images. CVIU 117(9), 966–989 (2013)
20. Tran, P.V.: A fully convolutional neural network for cardiac segmentation in short-axis mri. preprint arXiv:1604.00494 (2016)
21. Tulder, G. van, and Bruijne, M. de: Why does synthesized data improve multi-sequence classification? In: MICCAI, pp. 531–538 (2015)
22. Van Nguyen, H., Zhou, K., and Vemulapalli, R.: Cross-domain synthesis of medical images using efficient location-sensitive deep network. In: MICCAI, pp. 677–684 (2015)
23. Vemulapalli, R., Van Nguyen, H., and Kevin Zhou, S.: Unsupervised cross-modal synthesis of subject-specific scans. In: IEEE ICCV, pp. 630–638 (2015)
24. Zhou, Y., Giffard-Roisin, S., De Craene, M., D'hooge, J., Alessandrini, M., Friboulet, D., Sermesant, M., Bernard, O.: A Framework for the Generation of Realistic Synthetic Cardiac Ultrasound and Magnetic Resonance Imaging Sequences from the same Virtual Patients. IEEE TMI (2017)
25. Zhu, J.: Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. preprint arXiv:1703.10593 (2017)
26. Zhuang, X., Rhode, K.S., Razavi, R.S., Hawkes, D.J., and Ourselin, S.: A registration-based propagation framework for automatic whole heart segmentation of cardiac MRI. IEEE TMI 29(9), 1612–1625 (2010)
27. Zhuang, X., and Shen, J.: Multi-scale patch and multi-modality atlases for whole heart segmentation of MRI. Medical image analysis 31, 77–87 (2016)