UNIVERSITY OF LEEDS

This is a repository copy of *Unsupervised Standard Plane Synthesis in Population Cine MRI via Cycle-Consistent Adversarial Networks*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/160599/

Version: Accepted Version

Proceedings Paper:
Zhang, L, Pereañez, M, Bowles, C et al. (4 more authors) (2019) Unsupervised Standard Plane Synthesis in Population Cine MRI via Cycle-Consistent Adversarial Networks. In: Lecture Notes in Computer Science. Medical Image Computing and Computer Assisted Intervention – MICCAI 2019, 13-17 Oct 2019, Shenzhen, China. Springer Verlag , pp. 660-668. ISBN 978-3-030-32244-1

https://doi.org/10.1007/978-3-030-32245-8_73

© Springer Nature Switzerland AG 2019. This is an author produced version of a conference paper published in Lecture Notes in Computer Science. Uploaded in accordance with the publisher's self-archiving policy.

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Unsupervised Standard Plane Synthesis in Population Cine MRI via Cycle-Consistent Adversarial Networks

Le Zhang[1,2(✉)], Marco Pereañez[3], Christopher Bowles[3], Stefan K. Piechnik[4], Stefan Neubauer[4], Steffen E. Petersen[5], Alejandro F. Frangi[3]

[1]Centre for Computational Imaging and Simulation Technologies in Biomedicine, Department of Electronic and Electrical Engineering, University of Sheffield, UK.
le.zhang@sheffield.ac.uk
[2]Queen Square Institute of Neurology, University College London, UK.
[3]Centre for Computational Imaging and Simulation Technologies in Biomedicine, School of Computing and School of Medicine, University of Leeds, UK.
[4]Oxford Center for Clinical Magnetic Resonance Research (OCMR), Division of Cardiovascular Medicine, University of Oxford, John Radcliffe Hospital, UK.
[5]William Harvey Research Institute, Queen Mary University of London, Barts Heart Centre, Barts Health NHS Trust, UK.

Abstract. In clinical studies or population imaging settings, cardiac magnetic resonance (CMR) images may suffer from artifacts due to variability in the breath-hold position adopted by the patient during the scan. Consistent orientation of image planes with respect to the cardiac ventricles in CMR sequences forms a crucial step in the assessment of cardiac function via parameters such as the Ejection Fraction (EF) and Cardiac Output (CO) of both ventricles, which are the most immediate indicators of normal/abnormal cardiac function. In this paper, we present a novel unsupervised approach for the realistic transformation of acquired CMR images to a standard orientation using Cycle-Consistent Adversarial Networks (Cycle-GANs). We tackle this challenge by splitting the problem into two principal subtasks. First, we consider a bidirectional generator mapping between the re-oriented image and the original, hence allowing direct comparison to the input image without the need to resort to paired training data. Second, we devise a novel loss function incorporating intensity and orientation terms, and aims to produce images of high perceptual quality. Extensive experiments conducted on the CMR images in the UK Biobank dataset demonstrate that the images rendered by our model can improve the accuracy of the image derived cardiac parameters.

Keywords: Deep Learning · Cycle-Consistent Adversarial Networks · Cardiac Orientation · Ventricular Volume · MRI · Population Imaging.

1 Introduction

Cardiac anatomy and function are widely used in diagnosis and monitoring of disease progression in cardiology, with CMR imaging arguably being one of the

most wide-spread techniques for clinical diagnostic imaging of the heart. CMR requires a carefully selected and consistent orientation of short-axis (SAX) image planes with respect to the cardiac ventricles, particularly the basal slice (BS) and apical slice (AS) plane which contain key anatomical structures [10]. If the plane orientation deviates significantly from expected values, local image structure may change enough to cause subsequent image feature-based algorithms to fail in localizing key features required for further morphological and functional analysis. However, it is challenging and time-consuming, even for experienced MRI operators, to manually find the correct imaging plane, particularly when it is subject to subsequent patient movement. The task is highly operator-dependent and requires a great amount of expertise. With the advent of CMRI, 2D SAX slices can be acquired quickly with little training. But the problem of locating the standard planes required for diagnostically important biometric measurements remains. There is a strong need to develop automatic methods for 2D standard plane generation from existing 2D slices to improve clinical workflow efficiency.

Image generation is a hot topic which has achieved great success on many vision tasks such as text-to-image generation [8] and image style transformation [5]. Conditional GAN [6] are more advanced image generators and more suitable for image translation tasks. It is developed by adding an input condition vector, which can include a vast amount of information, to the generator. Medical image synthesis is currently an emerging area of interest for applying the latest image generation techniques mentioned above. Zhou et al. [11] use a cycle consistency loss as a way of mapping back the initially rendered image to the original image to achieve unpaired image-to-image translation. They have shown impressive results in rendering new realistic images. Nie et al. [7] proposed a context-aware GAN by adding an image gradient difference term to the loss function of the generator, to retain the sharpness of the generated images. Dar et al. [1] utilized the image conditioned architectures Cycle-GAN and pix2pix to generate T1- from T2-weighted MR contrast and vice versa.

Inspired by this idea, we propose a fully unsupervised approach based on GANs that, given a SAX slice with incorrect plane orientation (IPO), automatically generates images under the correct orientation. To train our model using unlabeled data (*i.e.*, our training data consists of query slices and slices with correct orientation from other cardiac volumes), we propose a Cycle-GAN based architecture that aims at producing new images of high perceptual quality [2] by combining loss functions used in orientation transfer. The main contributions of the standard plane synthesis GAN (SPSGAN) are highlighted:

(1) A novel deep architecture is proposed for generating SAX slices with standard plane orientation. To achieve this, we devised a novel loss function computed over the images from a Cycle-GAN for cardiac orientation transfer.

(2) We propose a fully unsupervised strategy trained without paired training examples for image-to-image translation.

(3) This is the first paper to exploit deep learning method, especially GANs, for orientation based cardiac slice generation, which is an important step after quality control (QC) and before quantitative CMR analysis.

2 Methodology

2.1 Problem Formulation

To produce realistic standard orientation transformations of an input slice while retaining the intensity appearance, we use a single SAX slice as input and train a GAN model using an unsupervised approach. Formally, we seek to learn the mapping $(\mathbf{x}_t^i, \mathbf{f}_{\theta_o}, \mathbf{f}_{\gamma_o}) \rightarrow \mathbf{x}_o^i$ between an image $\mathbf{x}_t^i \in \mathbb{R}^{H\times W\times Z}$ in a volume with incorrect orientation $< \theta_t, \gamma_t >$ and the image $\mathbf{x}_o^i \in \mathbb{R}^{H\times W\times Z}$ of the same volume with the correct orientation $< \theta_o, \gamma_o >$. Orientations are represented by $< \theta, \gamma >$, where θ indicates the deflection angle in the xoy plane and γ indicates the deflection angle in the z direction of the 3D coordinate system. The subscripts o and t denote the correct and transformed orientations, respectively. The model is trained using an unsupervised approach with training samples $\{\mathbf{x}_o^i, \mathbf{x}_t^j\}_{i,j=1}^N$, which do not include the ground-truth image $\mathbf{x}_o^j$.

2.2 Unsupervised SPSGAN

Figure. 1 shows the structure of our SPSGAN model consisting of five main modules: (1) The real orientation features $[\mathbf{f}_\theta, \mathbf{f}_\gamma]$ are learned from different cardiac volumes and concatenated with the features in the generator to better generate an image with the desired orientation. (2) A generator $G(\mathbf{x}\,|(\mathbf{f}_\theta, \mathbf{f}_\gamma))$ that maps one given slice under an incorrect orientation to an output slice under the standard orientation with the same cardiac identity. G is used twice in our network, first to map the input image $\mathbf{x}_{tr}^i \rightarrow \mathbf{x}_{og}^i$ and then render the latter back to the initial orientation $\mathbf{x}_{og}^i \rightarrow \hat{\mathbf{x}}_{tg}^i$; (3) A regressor R responsible of estimating the slice orientation of a given image. Note that R is different from the pre-trained regression net for feature extraction in (1); (4) A discriminator D that tries to discriminate the generated and real images; (5) A loss function that is computed without ground-truth and aims to preserve the cardiac intensity. To address this challenge, we propose a novel loss function that enforces the intensity content similarity of $\mathbf{x}_{tr}^i$ and $\hat{\mathbf{x}}_{tg}^i$, and orientation similarity between $\mathbf{x}_{og}^i$ and $\mathbf{x}_{or}^j$. In the following, we describe in detail each of the five modules.

Orientation Feature Embedding: The orientations of all volumes are linearly distributed and categorized with two parameters $< \theta, \gamma >$. Since the loss of the orientation regressor can be fast and well converged during the training process, we can easily learn each orientation feature cluster from all the training data by k-means clustering, and compute the feature in the center of the cluster as a condition to generate images in desired orientation. The feature of each orientation in an image $\mathbf{x}_t^i \in \mathbb{R}^{H\times W\times Z}$ is represented as a probability density map $\mathbf{f}$ computed over the entire image domain as:

$$\mathbf{f} = \mathit{\Gamma}(series(\{\mathbf{x}_n\}_{n=1}^N)) \tag{1}$$

where the operator $\mathit{\Gamma}(\cdot)$ is to extract the feature of the input image $\mathbf{x}$ and N is the defined number of slices.

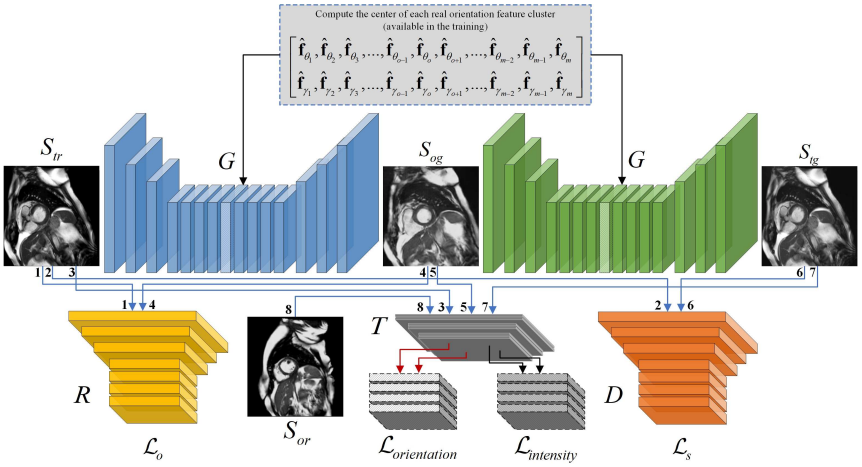

Fig. 1. The structure of our SPSGAN to generate the standard plane of the cardiac MRI in SAX view. Our model consists of five main components: a generator G, a discriminator D, an orientation regressor R, the transfer net T and the pretrained orientation features. S_{tr} is the original (IPO) image, S_{og} is the synthesized image, S_{tg} is the rendered back image, S_{or} is the image with correct orientation of different volume. Neither ground-truth image is considered.

Generator: Given an input image $\mathbf{x}$ with incorrect orientation, the generator $G(\mathbf{x}\,|(\mathbf{f}_\theta, \mathbf{f}_\gamma))$ aims to render $\mathbf{x}$ in a standard orientation with $< \theta_o, \gamma_o >$. To condition the generator with the orientation features we consider the concatenation $(\mathbf{x}, \mathbf{f}_\theta, \mathbf{f}_\gamma) \in \mathbb{R}^{H\times W\times Z}$ and feed this into a feedforward network, which generates output images of the same size as $\mathbf{x}$. To achieve improved image-to-image translation results, we adopt the network variation from [4] to construct the generator.

Image Discriminator: We adopt the PatchGAN [3] network as the discriminator $D(\mathbf{x})$, which maps from the input image $\mathbf{x}$ to a matrix $Y_s \in \mathbb{R}^{26\times 26}$. The discriminator then classifies each 26×26 patch in an image as real or fake. Since a smaller PatchGAN can generate high perceptual quality images with fewer parameters and less time [3], we run the discriminator across the image in a convolutional manner and average all responses to provide the final output D.

Orientation Regressor: D distinguishes the generated samples from the real images. We simultaneously use an orientation regressor R to regress the inferred slice with correct orientations. R is implemented with the ResNet architecture described in [11].

2.3 Optimisation

We have three terms to be optimized for the full loss function. 1) A generative adversarial loss that enforces the distribution of the generated images to be similar to that of the training images. 2) An orientation regression loss that

enforces the orientation of the generated images to be similar to the standard orientation. 3) The transfer loss that preserves the cardiac identity between the generated and the input images. Next, we will describe each of these in detail.

Generative Adversarial Loss: To optimize the parameters of generator G and learn the distribution of the training data, we perform a standard *minmax* game between G and discriminator D. G and D are jointly trained with the objective function $\mathcal{L}_s(G, D, \mathbf{x}, \mathbf{f}_\theta, \mathbf{f}_\gamma)$ where D tries to maximize the probability of correctly classifying original and rendered images while G tries to fool D.

$$\mathcal{L}_s(G, D, \mathbf{x}, \mathbf{f}_\theta, \mathbf{f}_\gamma) = E[\log D(\mathbf{x})] + E[\log(1 - D(G(\mathbf{x}\,|(\mathbf{f}_\theta, \mathbf{f}_\gamma)\,)] \tag{2}$$

Orientation Regression Loss: G must not only maximise the loss of D, but also must reduce the error produced by the orientation regressor R. In this way, while learning to produce realistic samples, G also learns how to generate images with the standard orientation $< \theta, \gamma >$. This loss is defined by:

$$\mathcal{L}_o(G, R, \mathbf{x}, \mathbf{f}_\theta, \mathbf{f}_\gamma) = \|R(G(\mathbf{x}\,|(\mathbf{f}_\theta, \mathbf{f}_\gamma\,)))- < \mathbf{f}_\theta, \mathbf{f}_\gamma >\|_2^2 \tag{3}$$

Transfer Loss: With the two previously defined losses $\mathcal{L}_s$ and $\mathcal{L}_o$, G is enforced to generate realistic slices with correct orientation. However, in the absence of ground-truth supervision, there is no constraint to ensure appearance identity. We derive inspiration from the previously introduced content-style loss to maintain high perception quality in image style transfer [2]. The loss mainly consists of two parts, one retains intensity similarity and the other transfers orientation similarity. Inspired by this idea, we define two sub-losses to maintain the identity between the input slice $\mathbf{x}^i_{tr}$ and the rendered slice $\mathbf{x}^i_{og}$.

For the intensity term, we define that G should be able to render-back the initial slice $\mathbf{x}^i_{tr}$ given the generated slice $\mathbf{x}^i_{og}$ and the original orientation features $< \mathbf{f}_{\theta_t}, \mathbf{f}_{\gamma_t} >$, that is $\hat{\mathbf{x}}^i_{tg} \approx \mathbf{x}^i_{tr}$, where $\hat{\mathbf{x}}^i_{tg} = G(G(\mathbf{x}^i_{tr}\,|(\mathbf{f}_{\theta_o}, \mathbf{f}_{\gamma_o}))\,|(\mathbf{f}_{\theta_t}, \mathbf{f}_{\gamma_t}))$. However, it is difficult to handle high frequency details by directly comparing $\mathbf{x}^i_{tr}$ and $\hat{\mathbf{x}}^i_{tg}$ using Patch-GAN at a pixel level, which will lead to overly-smoothed images. Instead, we compare them based on their intensity content. Formally, we define the intensity loss to be:

$$\mathcal{L}_{intensity} = \left\|T_l(\mathbf{x}^i_{tr}) - T_l(\mathbf{x}^i_{tg})\right\|_2^2 \tag{4}$$

where $T_l(\cdot)$ represents the feature representation at the l^{th} layer of the network.

In order to transfer the standard orientation information from the real slice to the synthesized one, we take over the spatial extent of the feature maps to design the feature space for capturing texture information. Previous work [2] implements this by computing the Gram matrix $\mathbf{M}^l \in \mathbb{R}^{U \times U}$, where $\mathbf{M}^l$ is the inner product between the vectorised feature maps of $\mathbf{x}^i_{og}$. The orientation loss is then computed as the mean square error between visible pairs of Gram matrices of the same joint in both images $\mathbf{x}^i_{og}$ and $\mathbf{x}^j_{or}$:

$$\mathcal{L}_{orientation} = \frac{1}{L}\sum_{l=0}^{L}\left(\frac{\mathbf{M}^{i,l}_{og} - \mathbf{M}^{j,l}_{or}}{UV}\right)^2 \tag{5}$$

where $\mathbf{M}_{og}^{i,l}$ and $\mathbf{M}_{or}^{j,l}$ are the orientation representations in the layer l of the generated and real image with standard orientation, respectively. In layer l, there is U_l feature maps each of size V_l, where V_l is the height times the width of the feature map. Finally, we define the transfer loss as the weighted sum of the intensity and orientation losses:

$$\mathcal{L}_{TS} = \mathcal{L}_{content}(T, \mathbf{x}_{tr}^{i}, \hat{\mathbf{x}}_{tg}^{i}) + \lambda \mathcal{L}_{orientation}(T, \mathbf{x}_{tr}^{i}, \mathbf{x}_{og}^{i}, \mathbf{x}_{or}^{j}) \tag{6}$$

where the parameter λ controls the relative importance or the two components.

Full Loss: We take the full loss as a linear combination of all previous loss terms:

$$\mathcal{L}_{SPS} = \arg\min_{G} \max_{D,R,T} \{\mathcal{L}_s(G, D, \mathbf{x}, \mathbf{f}_\theta, \mathbf{f}_\gamma) + \alpha \mathcal{L}_o(G, R, \mathbf{x}, \mathbf{f}_\theta, \mathbf{f}_\gamma) + \mathcal{L}_{TS}\} \tag{7}$$

where α is the weighting factors for image adversarial and orientation regression loss ($\alpha = 400$ and $\lambda = 0.2$ in this work).

3 Experiments and Analysis

Materials and Evaluation Metrics: We used a series with the first 5,000 CMR subjects available from the UK Biobank (UKBB) imaging resource, with each volumetric sequence containing about 50 cardiac phases. Based on analysis of the in-plane orientation angle distribution for the 5,000 subjects for which manual segmentations are available (and therefore θ, γ can be computed), we found that θ has the median value of 132.8° and standard deviation of 8.0°, while γ has the median value of 7.1° with standard deviation of 3.9°. Among them, there are 302 cases under standard cardiac orientations ($\theta = 135°, \gamma = 0°$). The set of orientation labels were chosen from these realistic distributions and trained in a regression net to obtain the real orientation features.

Since our model is trained using the unsupervised approach and there is no ground-truth for the test images, we need to generate the slices with correct orientation as reference samples to evaluate the synthetic images. The reference images are resampled from the interpolated 3D cardiac volumes by Paraview[1]. The resampled slices are chosen with correct orientations and the same position (*i.e.*, the distance to base and apex) compared with the original images.

Experimental Settings: We verify the effectiveness of the unsupervised SPSGAN model through two groups of experiments. In the first experiment, the synthetic slice is evaluated against the reference image using rotation angles between the planes. Image similarity of the planes is also measured using peak signal-to-noise ratio (PSNR) and structural similarity (SSIM). In the second experiment, we evaluate synthetic slices against the corresponding reference slices on the tasks of LV segmentation and the measurement of cardiac function based on blood volumes. Four parameters are used for performance evaluation, including two commonly used indexes of the cardiac function derived from such

[1] https://www.paraview.org/

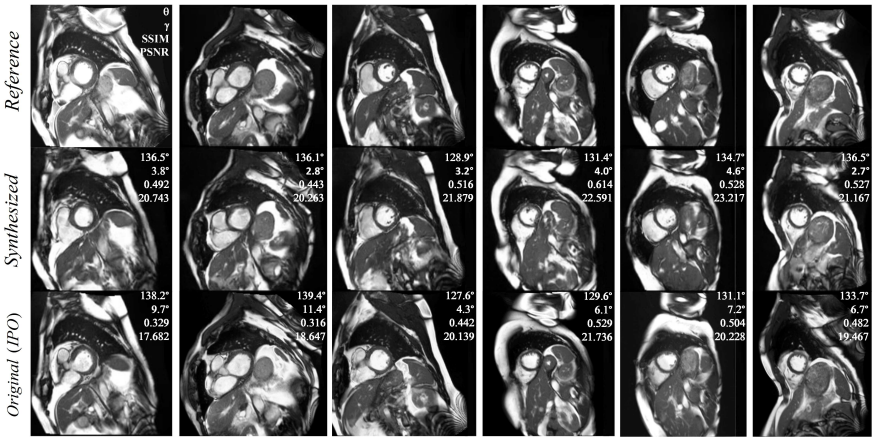

Fig. 2. Synthesized images by SPSGAN and corresponding original (IPO) images with orientation angles, PSNR and SSIM values, compared to the references.

Table 1. Effect of IPO on the ED, ES, SV and EF. Values are shown as mean ± standard deviations.

	Reference Image	Synthetic Image	Effect(%)	IPO Image	Effect(%)
LVEDV(ml)	159.6±32.7	151.5±34.9	-5.1%	142.9±31.5	-10.5%
LVESV(ml)	72.4±23.1	68.3±20.3	-5.7%	64.3±22.4	-11.2%
LVSV(ml)	87.2±17.6	83.2±18.4	-4.6%	78.6±17.9	-9.9%
LVEF(%)	54.6±0.08	54.9±0.09	+0.5%	55.0±0.08	+0.7%

volumes, stroke volume (SV) and EF, and similarly report differences between real and imputed image data.

Performance of Image Synthesis. We train the SPSGAN model using the 302 subjects with correct orientation and the same number of cases with incorrect orientations in UKBB, and test the model on another 100 subjects with incorrect orientations with comparisons to the corresponding resampled reference slices. Training images are only associated to the original slices with correct and incorrect orientations. No reference images are considered during training. Several typical images with real and synthetic slices are shown in Fig. 2. We can observe that our synthetic images show a slight difference from their corresponding original images, but similar to their corresponding reference images. This is because the local image structure in planes with different orientations will change. The orientation angles, SSIM and PSNR between synthetic, original and reference slices are also shown in Fig. 2. These results imply that our trained SPSGAN model is reasonable, and the synthetic CMR images have an acceptable representation in the standard planes.

Cardiac Functional Parameters Calculation. To assess the impact of synthetic images in real applications, such as the measurement of cardiac function based on blood volumes, we design an experiment to measure the differences between volumes derived from the reference volumes, synthetic volumes and original volumes with incorrect cardiac orientation. The experimental results across four different cardiac parameters using the LV segmentation method described in [9] are reported in Table 1. For this experiment, we compute blood pool volumes at the End-diastolic (ED) and End-systolic (ES) phases, and from these, we obtain SV and EF. The average volumes and indexes are computed across the sample, comparing the reference volumes, synthetic volumes and incomplete volumes. Table 1 shows that the incorrect plane orientation reduces ED and ES volumes by an average of 11%. In contrast, the synthetic images provide values which are much closer than the reference values, with only 5.1% and 5.7% reductions in volume at ED and ES phases. These results clearly demonstrate synthetic images generated by SPSGAN model convey relevant information and possess clinical utility.

4 Conclusion

We have presented a novel approach for generating cardiac cine MRI slices under a standard ventricle plane orientation using a GAN model that can be trained using a fully unsupervised approach. Finding the correct standard plane is highly operator-dependent and requires a great amount of expertise. To tackle this challenge, we proposed a fully unsupervised framework that aims to transfer the plane orientation and retaining the cardiac intensity of the original image without depending on the corresponding ground-truth. Extensive experimental results showed that our model could achieve satisfactory performance in standard cardiac slice generation compared to other methods. In the future, we plan to further apply our approach to other datasets and to different modalities for which supervision is not possible.

References

1. Dar, S.U., Yurt, M., Karacan, L., Erdem, A., Erdem, E., Çukur, T.: Image synthesis in multi-contrast MRI with conditional generative adversarial networks. IEEE transactions on medical imaging (2019)
2. Gatys, L.A., Ecker, A.S., Bethge, M.: Image style transfer using convolutional neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2414–2423 (2016)
3. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks pp. 1125–1134 (2017)
4. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: European conference on computer vision. pp. 694–711. Springer (2016)
5. Liu, M.Y., Tuzel, O.: Coupled generative adversarial networks. In: Advances in Neural Information Processing Systems. pp. 469–477 (2016)

6. Mirza, M., Osindero, S.: Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784 (2014)
7. Nie, D., Trullo, R., et al: Medical image synthesis with context-aware generative adversarial networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 417–425. Springer (2017)
8. Reed, S., Akata, Z., Yan, X., Logeswaran, L., Schiele, B., Lee, H.: Generative adversarial text to image synthesis. arXiv preprint arXiv:1605.05396 (2016)
9. Tran, P.V.: A fully convolutional neural network for cardiac segmentation in short-axis MRI. arXiv preprint arXiv:1604.00494 (2016)
10. Zhang, L., Gooya, A., Pereaez, M., Dong, B., Piechnik, S.K., Neubauer, S., Petersen, S.E., Frangi, A.F.: Automatic assessment of full left ventricular coverage in cardiac cine magnetic resonance imaging with fisher-discriminative 3-D CNN. IEEE Transactions on Biomedical Engineering 66(7), 1975–1986 (July 2019)
11. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision. pp. 2223–2232 (2017)