Unsupervised Medical Image Translation Using Cycle-MedGAN

Karim Armanious[1,2], Chenming Jiang[1], Sherif Abdulatif[1], Thomas Küstner[1,2,3], Sergios Gatidis[2], Bin Yang[1]
[1]University of Stuttgart , Institute of Signal Processing and System Theory, Stuttgart, Germany
[2]University of Tübingen, Department of Radiology, Tübingen, Germany
[3]King's College London, Biomedical Engineering Department, London, England

***Abstract*—Image-to-image translation is a new field in computer vision with multiple potential applications in the medical domain. However, for supervised image translation frameworks, co-registered datasets, paired in a pixel-wise sense, are required. This is often difficult to acquire in realistic medical scenarios. On the other hand, unsupervised translation frameworks often result in blurred translated images with unrealistic details. In this work, we propose a new unsupervised translation framework which is titled Cycle-MedGAN. The proposed framework utilizes new non-adversarial cycle losses which direct the framework to minimize the textural and perceptual discrepancies in the translated images. Qualitative and quantitative comparisons against other unsupervised translation approaches demonstrate the performance of the proposed framework for PET-CT translation and MR motion correction.**

***Index Terms*—Medical image translation, Unsupervised Learning, PET-CT, GANs, Motion Correction**

I. Introduction

In recent years, the machine learning community has achieved tremendous leaps in performance. This owes to increasingly available computational resources and open-source access to large datasets. From another perspective, radiological scans are vital tools in modern medicine. They enable diagnostics, disease tracking and patient treatment monitoring. This has led to the utilization of recent advances in computer vision, especially Deep Convolutional Neural Networks (DCNNs), in the field of medical image analysis. For example, DCNNs have been adapted for lesion classification in Magnetic Resonance Images (MRI) [1], 3D image segmentation [2] and anomaly detection [3] among other applications [4], [5].

A branch of deep learning is generative models which are utilized for dataset generation and augmentation. Amongst them, Generative Adversarial Networks (GANs) [6] are the prominent choice, with a large body of research focusing on theoretical and architectural analysis [7], [8]. In 2016, the pix2pix framework, a supervised GAN-based framework, has introduced the task of image-to-image translation from a source domain image, e.g. a day-time image, to a corresponding target domain image, e.g. a night-time image, provided that both domains have the same underlying structure [9]. This task has been adapted to the field of medical image analysis by using pix2pix for applications such as low-dose Computed Tomograph (CT) denoising [10], Positron Emission-computed Tomography (PET) to MR translation [11], splenomegaly segmentation [12] and MR to CT translation [13]. Other specialized architectures have been introduced for tasks such as compressed sensing MR reconstruction [14] and retinal image super-resolution [15]. Recently, we proposed MedGAN as a new framework for image translation [16]. It extends pix2pix with a cascaded U-net generator architecture [17] and additional non-adversarial losses, such as perceptual [18] and style transfer loss functions [19]. Since then it has been applied to tasks such as PET denoising [16], MR motion artifacts correction [20] and medical image in-painting [21].

However, these methods are supervised. In other words, training such models requires paired datasets where the images from the source domain are paired in a pixel-wise sense with the corresponding images in the target domain. Nevertheless, the acquisition of such paired datasets in real-life situations is often challenging. This is due to difficulties in obtaining co-registered cross-modality data from different scanners and acquisition sequences or multi-modal data for some organs such as the heart due to technical challenges or the required extensive planning and acquisition. Consequently, unsupervised image translation techniques, which are trainable with no paired examples but with image samples from both domains, are especially important for medical image translation.

Several methods for unsupervised image-to-image translation have been developed. UNIT is an unsupervised translation framework which maps the input-target images into a common latent space using a pair Variational Auto Encoder-GANs (VAE-GANs) before reconstruction in the desired image domain [22]. It has been utilized in the medical domain for the translation of T1-weighted and T2-weighted MR scans [23]. Cycle-GAN is another unsupervised translation approach which is based on the combination of adversarial losses and the pixel-wise cycle-consistency loss [24]. It has been adapted for medical translation tasks such as CT to MR bidirectional-translation [25] and CT denoising [26]. Other unsupervised frameworks exist with an overview available in [27].

In this work, we introduce a new framework for unsupervised medical image translation titled Cycle-MedGAN. This work expands the Cycle-GAN framework by introducing two new non-adversarial loss functions analogous to the perceptual and style transfer losses utilized in MedGAN. However, unlike MedGAN, the calculation of such losses does not require any explicit pairing of the input datasets during training. The training procedure is unsupervised using unpaired data, while validation is conducted on paired datasets.To validate the performance of the proposed framework, qualitative and quantitative comparisons against unsupervised frameworks, such as Cycle-GAN and UNIT, are conducted. The comparisons are carried out on the two medical tasks, MR motion artifact correction and PET to CT translation.

978-9-0827-9703-9/19/$31.00

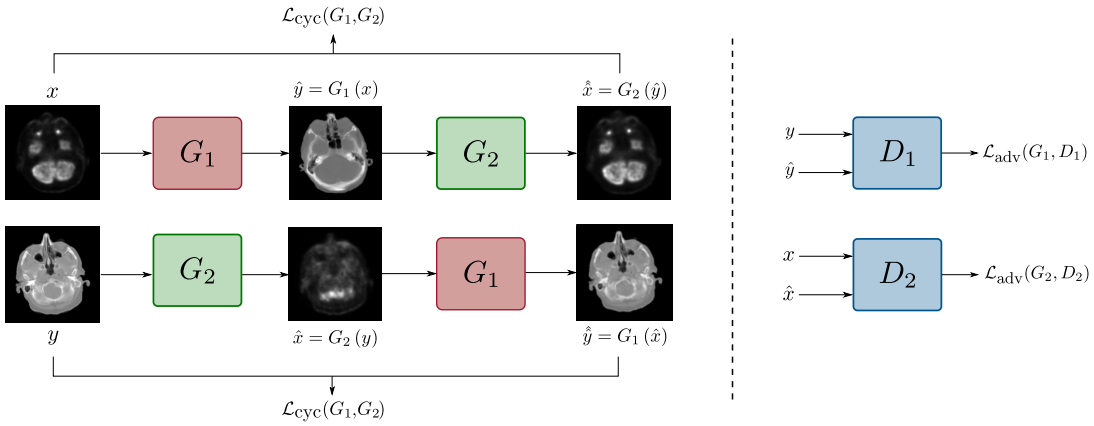

Fig. 1: An overview of the Cycle-GAN framework for unpaired image translation. x and y are unpaired images randomly sampled from their respective domains.

II. Method

The proposed Cycle-MedGAN framework is based on the traditional Cycle-GAN framework with the inclusion of new non-adversarial losses. The baseline Cycle-GAN model is illustrated in Fig. 1.

A. The Cycle-GAN Framework

Cycle-GAN is an unsupervised framework which allows bidirectional translation between the source domain X, e.g. PET images, and the target domain Y, e.g. CT images. It consists of two mapping functions $G_1 : X \rightarrow Y$ and $G_2 : Y \rightarrow X$, where G_1 and G_2 are two generator networks parametrized using DCNNs. Each of the generator networks is trained adversarially using a corresponding discriminator network, D_1 and D_2. For illustration, the first generator network G_1 receives as input a source domain image, $x \in X$, and outputs a synthetic translation, $\hat{y} = G_1(x)$. D_1 receives as input both the synthetic output $\hat{y}$ and an unpaired image randomly sampled from the desired target domain, $y \in Y$. The two networks, G_1 and D_1, are pitted against each other in competition, where D_1 acts as a binary classifier attempting to distinguish between the translated samples and the target domain samples. On the other hand, G_1 attempts to improve the quality of the translated output, thus deceiving the discriminator. This training procedure is formulated as a min-max optimization task over the adversarial loss function $\mathcal{L}_{\text{adv}}(G_1, D_1)$:

$$\min_{G_1}\max_{D_1}\mathcal{L}_{\text{adv}}(G_1, D_1) = \mathbb{E}_y\left[\log D_1(y)\right] + \mathbb{E}_x\left[\log\left(1 - D_1\left(G_1(x)\right)\right)\right] \tag{1}$$

and $\mathcal{L}_{\text{adv}}(G_2, D_2)$ is the analogous loss function for the second pair of networks, G_2 and D_2, formed by replacing the input images as y and the translated outputs as $\hat{x}$.

Training the framework merely with the adversarial losses is not sufficient since it may lead to mode collapse, where a set of different input images are mapped into a single image in the target domain [24]. Therefore, an additional constraint regularizing the mapping functions is essential. This is achieved by Cycle-GAN which enforces the two mapping functions, G_1 and G_2, to be cycle-consistent with each other. In other words the two generator networks should invert each other such that $\hat{\hat{x}} = G_2(G_1(x)) \approx x$ and $\hat{\hat{y}} = G_1(G_2(y)) \approx y$. This behaviour can be incentivized by using the pixel-wise cycle-consistency loss for both generators:

$$\mathcal{L}_{\text{cyc}}(G_1, G_2) = \mathbb{E}_x\left[\|x - G_2\left(G_1(x)\right)\|_1\right] + \mathbb{E}_y\left[\|y - G_1\left(G_2(y)\right)\|_1\right] \tag{2}$$

B. Non-Adversarial Cycle Losses

Cycle-GAN relies on the cycle-consistency loss to avoid mismatches which could occur due to unsupervised training using unpaired images. However, it has been discussed in the literature that pixel-wise losses fail to capture the perceptual aspect of human judgement on image quality [28]. Thus, when used in translation tasks, they often lead to results which lack sharpness and fine-detailed structures [18], [19]. To circumvent this issue, feature-based loss functions were introduced as additional constraints to enhance the quality of translated output quality. For instance, the MedGAN framework utilized a combination of perceptual and style transfer losses [16]. However, for unsupervised image translation, the utilization of such loss functions is not viable. An unpaired translation model cannot be trained by penalizing the feature-based deviation of the translated image from the unknown ground truth image.

In this work, we propose an adaptation of the above feature-based loss functions for the task of unsupervised image translation. The penalized deviation is instead between the input images, x or y, and the cycle-reconstructed images, $\hat{\hat{x}}$ or $\hat{\hat{y}}$. This process is illustrated in Fig. 2. The first proposed loss function is the cycle-perceptual loss, $\mathcal{L}_{\text{cPercep}}$. Analogous to the perceptual loss introduced in [16], [18], this loss aims at minimizing the perceptual discrepancies and enhancing the global consistency of the output images. This is achieved by extracting intermediate feature maps, using a pre-trained feature extractor network, for both the input and the cycle-reconstructed images. The cycle-perceptual loss then calculated as the mean absolute error (MAE) between the extracted feature maps for both generators:

$$\mathcal{L}_{\text{cPercep}} = \sum_{i=0}^{L}\lambda_{cp,i}\left(\|F_i\left(x\right) - F_i\left(\hat{\hat{x}}\right)\|_1 + \|F_i\left(y\right) - F_i\left(\hat{\hat{y}}\right)\|_1\right) \tag{3}$$

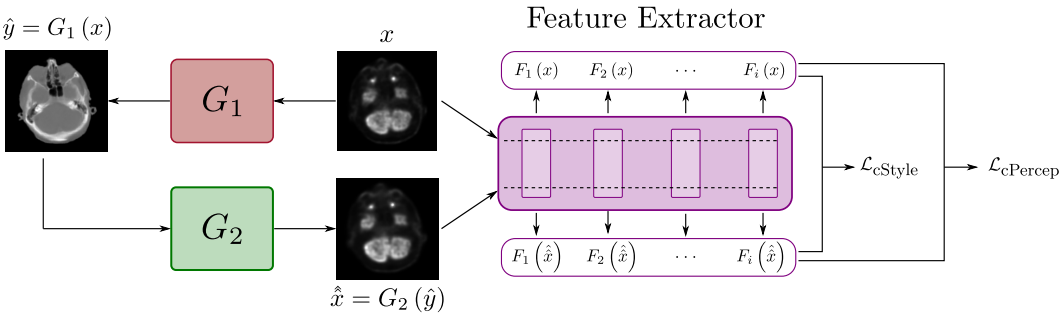

Fig. 2: An illustration of the proposed cycle non-adversarial loss functions calculated using a pre-trained feature extractor.

where F_i and F_j indicate the extracted feature maps from the i^{th} layer of the feature extractor network. L is the total number of layers, $\lambda_{cp,i}$ is the weight given to each layer.

The second proposed loss function is the cycle-style loss, which is typically utilized for style transfer applications [19]. This loss aims at matching the texture, style and fine details of the input images onto the cycle-reconstructed images. As a result, this motivates the generator architectures to produce more detailed translated outputs. The cycle-style loss is computed by first calculating the feature map correlations over the depth dimension. For G_1, this is represented by the Gram matrices, $Gr_i(x)$ and $Gr_i(\hat{x})$, whose elements are defined as:

$$Gr_i(x)_{m,n} = \frac{1}{h_i w_i d_i} \sum_{h=1}^{h_i} \sum_{w=1}^{w_i} F_i(x)_{h,w,m} F_i(x)_{h,w,n} \tag{4}$$

where h_i, w_i and d_i are the spatial height, width and depth of the extracted feature map of the i^{th} layer of the feature extractor network.

The style loss is then calculated as the weighted average of the squared Frobenius norm of the Gram matrices:

$$\begin{aligned} \mathcal{L}_{\text{cStyle}} = \sum_{i=1}^{L} \lambda_{cs,i} \frac{1}{4d_i^2} \Big(& \| Gr_i(x) - Gr_i(\hat{\hat{x}}) \|_F^2 \\ & + \| Gr_i(y) - Gr_i(\hat{\hat{y}}) \|_F^2 \Big) \end{aligned} \tag{5}$$

with $\lambda_{cs,i}$ is the weight given to the Gram matrices of the i^{th} layer.

For the Cycle-MedGAN framework, a combination of the adversarial, cycle-consistency, perceptual consistency and style consistency losses is utilized. The final min-max optimization task is given by:

$$\begin{aligned} \min_{G_1,G_2} \max_{D_1,D_2} \mathcal{L} = & \mathcal{L}_{\text{adv}}(G_1, D_1) + \mathcal{L}_{\text{adv}}(G_2, D_2) + \lambda_{cP} \mathcal{L}_{\text{cPercep}} \\ & + \lambda_{cyc} \mathcal{L}_{\text{cyc}}(G_1, G_2) + \lambda_{cS} \mathcal{L}_{\text{cStyle}} \end{aligned} \tag{6}$$

where λ_{cP}, λ_{cyc} and λ_{cS} are the weights given for the cycle-perceptual, cycle-consistency and cycle-perceptual losses respectively.

III. Datasets and Experiments

The Cycle-MedGAN framework was evaluated on two different tasks, PET to CT translation and the correction of motion artifacts in MR. For PET-CT translation, a dataset of the head region from 46 anonymized volunteers was acquired using a joint PET-CT scanner (Siemens Biograph mCT). For training, 1935 two-dimensional slices from 38 patients were utilized while the remaining 420 slices from 8 separate patients were used for validation. For the second application, T1-weighted MR data for the head region was acquired for 17 anonymized volunteers using a 3T MR scanner (Siemens Biograph mMT) with a fast spin echo sequence. Two different scans were acquired for each volunteer, one with voluntary rigid motion (head tilting) of the head and another under resting conditions [29]. Another 980 slices from 14 patients were used for training and 105 slices from the remaining 3 patients were used for validation. For both applications, the resolution of extracted data slices was re-sampled from their original resolutions to an isotropic voxel size of 1mm^3 and two-dimensional images of pixel dimensions 256×256 were extracted.

TABLE I: Quantitative comparison of unsupervised translation techniques

Model	(a) PET-CT translation			
	SSIM	PSNR(dB)	VIF	LPIPS
UNIT	0.8485	20.14	0.2057	0.6762
Cycle-GAN	0.8963	23.35	0.3831	0.2561
Cycle-MedGAN	**0.9115**	**24.08**	**0.4275**	**0.2233**
Model	**(b) MR motion correction**			
	SSIM	PSNR(dB)	MSE	UQI
UNIT	0.6914	18.64	0.1239	0.6953
Cycle-GAN	0.8011	22.39	0.3432	0.3282
Cycle-MedGAN	**0.8118**	**22.96**	**0.3513**	**0.3029**

Analogous to Cycle-GAN, random shuffling was applied between the different subjects in the collected datasets to ensure no explicit pairing between the source and target domains occur during the training procedure. The discriminator from a BiGAN network was utilized as the feature extractor network in all experimentations [30]. The BiGAN was pre-trained on a separate dataset of whole-body CT data for image reconstruction. This was conducted to extract plausible feature maps which would improve the quality of translated images.

To evaluate the performance of the proposed framework, qualitative and quantitative comparisons with other unsupervised translation techniques were carried out. More specifically, the UNIT framework [22] and the Cycle-GAN [24] were considered as performance baselines for the comparison. To ensure faithful representation of the baseline methods, verified open-source implementations were utilized along with the recommended hyper-parameters in the original publications [31], [32]. Besides the pre-trained feature extractor, the Cycle-

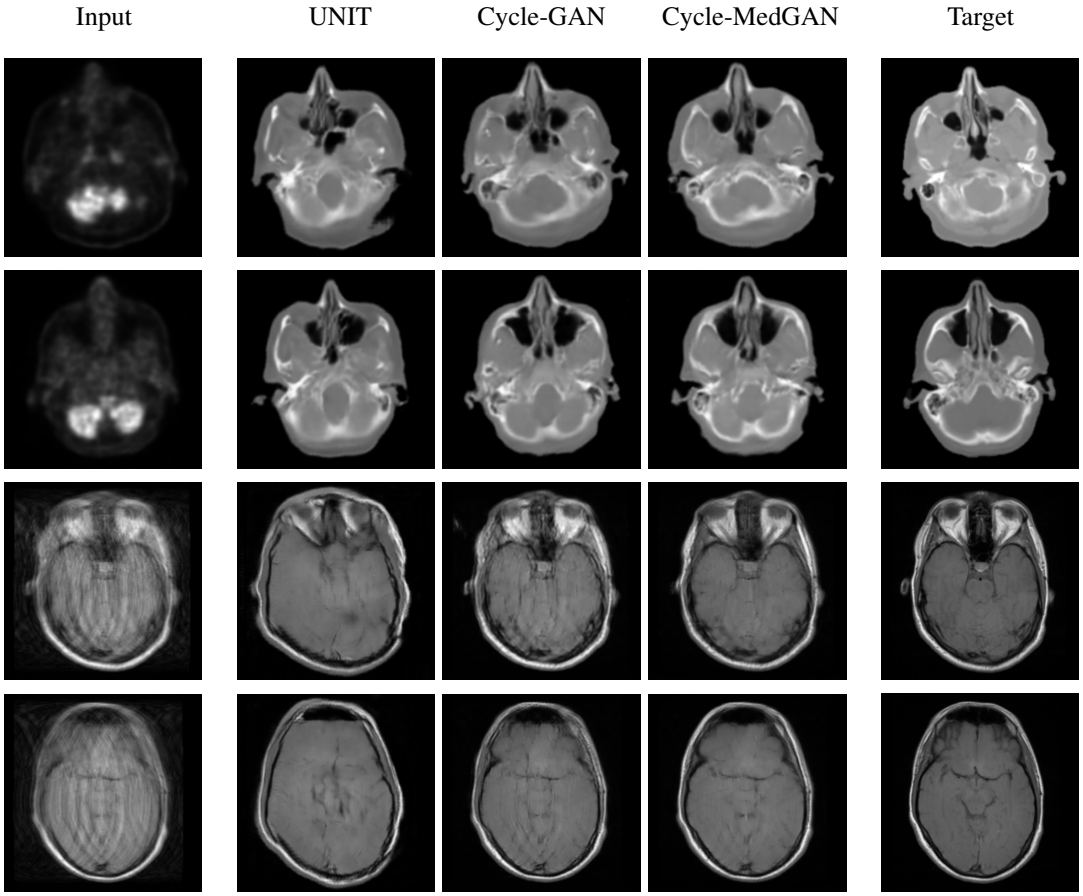

Fig. 3: Qualitative comparisons between the Cyclc-MedGAN framework and other unsupervised image translation techniques. The first two rows depict the task of PET to CT translation and last two rows illustrate the correction of MR motion artifacts.

MedGAN framework has an identical architecture as the Cycle-GAN framework which is described in details in [24]. The quantitative comparisons utilized the following metrics: Peak Signal to Noise Ratio (PSNR), Structural Similarity Index (SSIM) [33], Learned Perceptual Image Patch Similarity (LPIPS) [34] and Visual Information Fidelity (VIF) [35]. All models were trained using the ADAM optimizer and a batch size of 64. Training was for 50 epochs, lasting approximately 24 hours, using an NVIDIA Titan X GPU.

IV. Results and Discussion

The results of the proposed Cycle-MedGAN framework are presented in Table I and Fig. 3 in comparison with UNIT and Cycle-GAN. Qualitatively, the worse performance is exhibited by the UNIT framework. For both medical datasets in the comparative study, UNIT results in inhomogeneous global deformations as well as substantial blurs and distortion in the translated images. This is also reflected quantitatively, with UNIT resulting in the worst scores in Table I across the chosen metrics. In contrast, the resultant images produced by Cycle-GAN framework have a global structure which closely matches that of the target ground-truth images. However, finer details are not accurately translated by Cycle-GAN, such as the bone structures in the resultant CT images, and the motion blurring due to rigid motion in MR is not completely removed. The proposed Cycle-MedGAN framework builds upon the traditional Cycle-GAN architecture by introducing addition non-adversarial cycle losses to regulate the generator architectures. This results in an enhanced visual quality in the translated images (e.g. sharp edge delineation in MR images). The resultant CT images have noticeably sharper and more consistent bone structures compared to the other frameworks. Additionally, motion blurring due to rigid motion in MR is minimized. This enhancement in performance is analogously reflected quantitatively in Table I with Cycle-MedGAN surpassing the comparison baselines on all chosen metrics.

Despite the added level of quality by the Cycle-MedGAN framework, the proposed technique is not free from drawbacks. First, the resultant images by the Cycle-MedGAN framework potentially overlook important diagnostic information in the translation process. Thus, the framework is not intended for diagnostic applications but rather for post-processing tasks. An example of such tasks is using the synthetic CT images for PET attenuation correction or the calculation of organ volumes from motion corrupted MR images. Additionally, phase information contains vital information for motion correction in MR. Therefore, in future studies, we plan on expanding the framework with multi-channel three-dimensional inputs for complex-valued data. Furthermore, we plan to expand the comparison study to include performance on clinical applications and the effect of each individual non-adversarial

loss in comparison to different loss combinations.

V. Conclusion

Cycle-MedGAN is a new framework for unsupervised image translation. It builds upon the widely utilized CycleGAN framework with the additional utilization of new non-adversarial cycle loss functions, namely the cycle-perceptual loss and the cycle-style loss. The new loss functions use intermediate feature maps, extracted from a pre-trained feature extractor network, to direct the generator architectures to minimize perceptual and textural discrepancies in the results. Quantitative and qualitative comparisons with other unsupervised translation techniques indicate that the proposed framework enhances the translated outputs for the tasks of PET to CT translation and MR motion correction.

In the future, we plan to enhance the framework from the architectural aspect by incorporating three-dimensional complex-valued data. Additionally, the diagnostic performance of the framework will be investigated by experienced radiologists conducting subjective evaluations of the results.

References

[1] Q. Dou *et al.*, "Multilevel Contextual 3-D CNNs for False Positive Reduction in Pulmonary Nodule Detection," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 7, pp. 1558–1567, July 2017.

[2] K. Kamnitsas *et al.*, "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Medical Image Analysis*, vol. 36, pp. 61–78, 2017.

[3] T. Schlegl *et al.*, "Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery," in *Information Processing in Medical Imaging (IPMI)*, 2017, pp. 146–157.

[4] H. R. Roth *et al.*, "A New 2.5D Representation for Lymph Node Detection Using Random Sets of Deep Convolutional Neural Network Observations," in *Medical Image Computing and Computer-Assisted Interventio (MICCAI)*, 2014, pp. 520–527.

[5] H. Greenspan, B. van Ginneken, and R. M. Summers, "Guest Editorial Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting New Technique," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1153–1159, May 2016.

[6] I. J. Goodfellow *et al.*, "Generative Adversarial Networks," in *Conference on Neural Information Processing Systems (NIPS)*, 2014, pp. 2672–2680.

[7] T. Salimans *et al.*, "Improved Techniques for Training GANs," in *Conference on Neural Information Processing Systems (NIPS)*, 2016, pp. 2234–2242.

[8] I. Gulrajani *et al.*, "Improved Training of Wasserstein GANs," in *Conference on Neural Information Processing Systems (NIPS)*, 2017, pp. 5769–5779.

[9] P. Isola *et al.*, "Image-to-Image Translation with Conditional Adversarial Networks," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 5967–5976.

[10] J. M. Wolterink *et al.*, "Generative Adversarial Networks for Noise Reduction in Low-Dose CT," *IEEE Transactions on Medical Imaging*, vol. 36, no. 12, pp. 2536–2545, Dec 2017.

[11] H. Choi and D. S. Lee, "Generation of Structural MR Images from Amyloid PET: Application to MR-Less Quantification," *Journal of Nuclear Medicine*, vol. 59, pp. 1111–1117, 2018.

[12] Y. Huo *et al.*, "Splenomegaly segmentation using global convolutional kernels and conditional generative adversarial networks," in *Medical Imaging 2018: Image Processing*.

[13] H. E. M. *et al.*, "Generating synthetic CTs from magnetic resonance images using generative adversarial networks," *Medical Physics*, vol. 45, no. 8, pp. 3627–3636, 2018.

[14] T. M. Quan, T. Nguyen-Duc, and W. Jeong, "Compressed Sensing MRI Reconstruction Using a Generative Adversarial Network With a Cyclic Loss," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1488–1497, June 2018.

[15] D. Mahapatra *et al.*, "Image Super Resolution Using Generative Adversarial Networks and Local Saliency Maps for Retinal Image Analysis," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2017, pp. 382–390.

[16] K. Armanious *et al.*, "MedGAN: Medical Image Translation using GANs," http://arxiv.org/abs/1806.06397v1, 2018, arXiv preprint.

[17] S. Shah *et al.*, "Stacked U-Nets: A No-Frills Approach to Natural Image Segmentation," https://arxiv.org/abs/1804.10343, 2018, arXiv preprint.

[18] C. Wang *et al.*, "Perceptual Adversarial Networks for Image-to-Image Transformation," *IEEE Transactions on Image Processing*, vol. 27, 2018.

[19] J.Johnson, A. Alahi, and F. Li, "Perceptual Losses for Real-Time Style Transfer and Super-Resolution," 2016, pp. 694–711.

[20] K. Armanious *et al.*, "Retrospective correction of Rigid and Non-Rigid MR motion artifacts using GANs," https://arxiv.org/abs/1809.06276, 2019, accepted to IEEE International Symposium for Biomedical Images (ISBI).

[21] K. Armanious, Y. Mecky, S. Gatidis, and B. Yang, "Adversarial Inpainting of Medical Image Modalities," https://arxiv.org/abs/1810.06621, 2019, accepted to IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP).

[22] M. Liu, T. Breuel, and J. Kautz, "Unsupervised Image-to-Image Translation Networks," in *Conference on Neural Information Processing Systems (NIPS)*, 2017, pp. 700–708.

[23] P. Welander, S. Karlsson, and A. Eklund, "Generative Adversarial Networks for Image-to-Image Translation on Multi-Contrast MR Images - A Comparison of CycleGAN and UNIT," https://arxiv.org/abs/1806.07777, 2018, arXiv preprint.

[24] J. Zhu *et al.*, "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," in *IEEE International Conference on Computer Vision (ICCV)*, Oct 2017, pp. 2242–2251.

[25] C. Jin *et al.*, "Deep CT to MR Synthesis using Paired and Unpaired Data," https://arxiv.org/abs/1805.10790, 2018, arXiv preprint.

[26] E. Kang *et al.*, "Cycle-consistent adversarial denoising network for multiphase coronary CT angiography," *Medical Physics*, vol. 46, no. 2, pp. 550–562, 2019.

[27] X. Yi, E. Walia, and P. Babyn, "Generative Adversarial Network in Medical Imaging: A Review," https://arxiv.org/abs/1809.07294, 2018, arXiv preprint.

[28] D. Pathak *et al.*, "Context Encoders: Feature Learning by Inpainting," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2536–2544, 2016.

[29] T. Küstner *et al.*, "Automated reference-free detection of motion artifacts in magnetic resonance images," in *Magnetic Resonance Materials in Physics, Biology and Medicine*, vol. 31, 2018, pp. 243–256.

[30] J. Donahue, P. Krahenbühl, and T. Darrell, "Adversarial feature learning," in *International Conference on Learning Representations (ICLR)*, 2017.

[31] J. Kim, "UNIT implementation," https://github.com/taki0112/UNIT-Tensorflow.

[32] X. Hu, "Cycle-GAN implementation," https://github.com/xhujoy/CycleGAN-tensorflow.

[33] Z. Wang *et al.*, "Image quality assessment: from error visibility to structural similarity," in *IEEE Transactions on Image Processing*, vol. 13, 2004, pp. 600–612.

[34] R. Zhang *et al.*, "The Unreasonable Effectiveness of Deep Features as a Perceptual Metric," in *Conference on Computer Vision and Pattern Recognition*, 2018.

[35] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," in *IEEE Transactions on Image Processing*, vol. 15, 2006, pp. 430–444.