

# UCSF

## UC San Francisco Previously Published Works

### Title

Attention-Aware Discrimination for MR-to-CT Image Translation Using Cycle-Consistent Generative Adversarial Networks.

### Permalink

<https://escholarship.org/uc/item/847255k6>

### Journal

Radiology. Artificial intelligence, 2(2)

### ISSN

2638-6100

### Authors

Kearney, Vasant  
Ziemer, Benjamin P  
Perry, Alan  
[et al.](#)

### Publication Date

2020-03-01

### DOI

10.1148/ryai.2020190027

Peer reviewed

# Attention-Aware Discrimination for MR-to-CT Image Translation Using Cycle-Consistent Generative Adversarial Networks

Vasant Kearney, PhD • Benjamin P. Ziemer, PhD • Alan Perry, BS • Tianqi Wang, BS • Jason W. Chan, MD • Lijun Ma, PhD • Olivier Morin, PhD • Sue S. Yom, MD, PhD • Timothy D. Solberg, PhD

From the Department of Radiation Oncology, University of California, 1600 Divisadero St, San Francisco, CA 94115. Received March 3, 2019; revision requested April 15; revision received November 18; accepted November 25. Address correspondence to V.K. (e-mail: [vasantkearney@gmail.com](mailto:vasantkearney@gmail.com)).

Conflicts of interest are listed at the end of this article.

Radiology: Artificial Intelligence 2020; 2(2):e190027 • <https://doi.org/10.1148/ryai.2020190027> • Content codes: **IN** **CT** **MR**

**Purpose:** To suggest an attention-aware, cycle-consistent generative adversarial network (A-CycleGAN) enhanced with variational auto-encoding (VAE) as a superior alternative to current state-of-the-art MR-to-CT image translation methods.

**Materials and Methods:** An attention-gating mechanism is incorporated into a discriminator network to encourage a more parsimonious use of network parameters, whereas VAE enhancement enables deeper discrimination architectures without inhibiting model convergence. Findings from 60 patients with head, neck, and brain cancer were used to train and validate A-CycleGAN, and findings from 30 patients were used for the holdout test set and were used to report final evaluation metric results using mean absolute error (MAE) and peak signal-to-noise ratio (PSNR).

**Results:** A-CycleGAN achieved superior results compared with U-Net, a generative adversarial network (GAN), and a cycle-consistent GAN. The A-CycleGAN averages, 95% confidence intervals (CIs), and Wilcoxon signed-rank two-sided test statistics are shown for MAE (19.61 [95% CI: 18.83, 20.39],  $P = .0104$ ), structure similarity index metric (0.778 [95% CI: 0.758, 0.798],  $P = .0495$ ), and PSNR (62.35 [95% CI: 61.80, 62.90],  $P = .0571$ ).

**Conclusion:** A-CycleGANs were a superior alternative to state-of-the-art MR-to-CT image translation methods.

© RSNA, 2020

The use of CT is required in radiation therapy treatment planning (1). CT offers excellent anatomic localization information and provides the necessary electron density information needed for dose calculation (2). CT planning is increasingly complemented with MRI, as MRI enables superior soft-tissue visualization in many anatomic sites (3). However, nonrigid misalignment between CT and MR image sets introduces localization errors in target volumes and critical normal anatomy that are especially problematic for modern treatment modalities, such as intensity-modulated radiation therapy, which rely on accurately delineated anatomy (4–9). Furthermore, multiple imaging studies can be cost prohibitive and burdensome to the patient. In light of these challenges, MR-only treatment planning has become an attractive alternative. However, MR-to-CT image translation is technically challenging because of the nonlinear intensity and spatial translation between modalities.

In recent years, convolutional neural networks (CNNs) have been used for a broad range of volumetric prediction problems, including dose calculation, semantic segmentation, and synthetic CT generation from MR images (10–14). Conventional MR and CT image-synthesis CNNs learn to predict the most likely volume by minimizing voxel-to-voxel differences between MR and CT image pairs (15,16). This approach is prone to prediction degradation in the form of blurring and unsharpening due to anatomic misalignment (7,8).

Generative adversarial networks (GANs) provide an alternative to pure voxel-to-voxel learning by adding a discriminator CNN that encourages realistic predictions (17). Existing approaches use an encoder-decoder CNN called the generator to predict the translated image, which is passed into a discriminator CNN that classifies the quality of the translated image (18,19). The generator and discriminator compete to reach the Nash equilibrium, which is the minimax loss of the aggregate training protocol. However, this method still relies on voxel-to-voxel alignment because the objective function incorporates synthesis CNN loss. Because obtaining perfectly aligned MR and CT datasets is not possible, conventional GANs that rely on paired images have similar disadvantages as conventional CNNs.

Recently, a class of solutions has emerged that uses cycle-consistent GANs (CycleGANs) to solve image-to-image translation using unpaired images (20). These solutions solely rely on adversarial loss based on the discriminator, so they are not prone to translation degradation due to misalignment. Cycle consistency is introduced to encourage image translations that spatially correspond to their input images. Cycle consistency is accomplished by adding an additional GAN that predicts the original image based on the predicted translation. The difference between the reconstructed input image and the original image is added to the loss function to enforce spatial consistency between images. However, conventional unpaired CycleGANs rely

## Abbreviations

A-CycleGAN = attention-aware CycleGAN, CNN = convolutional neural network, CycleGAN = cycle-consistent GAN, GAN = generative adversarial network, MAE = mean absolute error, PSNR = peak signal-to-noise ratio, ReLU = rectified linear unit, SSIM = structure similarity index metric, VAE = variational autoencoding

## Summary

An architecture that uses a variational autoencoder-enhanced, attention-aware, cycle-consistent generative adversarial network (A-CycleGAN) for MR-to-CT image translation is described; this is the first time, to our knowledge, that an A-CycleGAN has been used to solve MR-to-CT image translation.

## Key Points

- An alternative to current state-of-the-art MR-to-CT image translation algorithms is suggested.
- Improved MR-to-CT image translation will facilitate MR-only radiation therapy treatment planning and MR-to-CT image fusion for diagnostic or therapeutic images.
- Successful implementation of this technique could help reduce the burden for patients who require MR and CT imaging in radiology and radiation oncology.

on the network's ability to distinguish a translation between domains, so they are limited by the network's ability to attend to specific anatomy (21).

An attention mechanism has been proposed to help the discriminator attend to specific anatomy within an image by selectively enhancing portions of the network during training (22). However, attention-gated classification networks are difficult to train (23). Recently, variational autoencoding (VAE) has been used to facilitate model convergence and improve the generalization of CNNs (24). This study suggests an attention-aware CycleGAN (A-CycleGAN) with VAE enhancement as an alternative to current state-of-the-art MR-to-CT image translation algorithms.

## Materials and Methods

### Data

Findings from 90 patients with head, neck, and brain cancer who previously received radiation therapy were used to train, validate, and test A-CycleGAN (institutional review board identifier 14-15452). To mitigate multiple-hypothesis testing, this study followed Kaggle-style competition rules, in which studies from 60 patients (4138 axial slices) were used to train and validate the model. Studies from an additional 30 holdout-test patients (1422 axial slices) were deconstructed into axial slices for prediction and then reconstructed into three-dimensional volumes to report final test scores. All evaluation metrics and statistics were conducted using three-dimensional image volumes as opposed to separate slices. For each patient, a single CT and T1-weighted MR image set was used. All images were rescaled to  $1 \times 1 \times 2$  mm in the left-right, anterior-posterior, and superior-inferior directions, respec-

tively. CT images were acquired helically using 120 kVp and 450–720 mAs and using a SOMATOM Sensation CT scanner (Siemens Medical Solutions, Ann Arbor, Mich). T1-weighted MR images were acquired using a repetition time of 8.48 msec, an echo time of 10.584 msec, and a flip angle of  $111^\circ$  using a Discovery MR750 T3 MRI scanner (GE Medical Systems, Chicago, Ill). All images were normalized prior to training by subtracting the mean and dividing by the standard deviation.

### Attention-Aware Discrimination

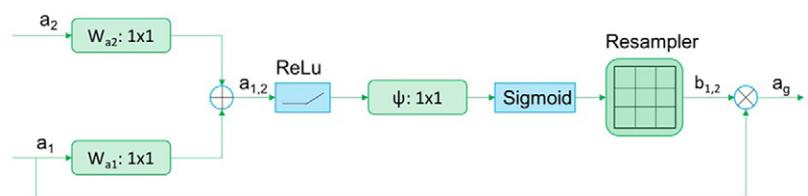
The discriminators use an attention-gated  $70 \times 70$  pixels "PatchGAN" classifier, which selectively captures local style statistics (20,22). The attention mechanism imposed on our discriminator networks encourages a more parsimonious use of image information and aids in generalization over changes in data distribution, enforcing compatibility between local feature vectors extracted at intermediate stages in the CNN pipeline and the final output function (23,25). Additive self-attention gates were used to modulate feature responses from the intermediate multiscale stages of the network (26,27). Each intermediate multiscale output is modulated by adding the output of the final multiscale stage ( $a_1$ ) to the output of each intermediate stage ( $a_2$ ). The combined activations ( $a_{1,2}$ ) are rectified linear unit (ReLU)-activated and passed through a  $1 \times 1$  channel-wise convolutional layer before being batch normalized and sigmoidally activated to form  $b_{1,2}$ . The  $a_1$  is then multiplied by  $b_{1,2}$  to form  $a_g$ . The  $a_g$  from each multiscale stage is concatenated to the final aggregated output stage. A schematic of the attention-gating mechanism is shown in Figure 1.

Training attention-gated classifier networks is nontrivial because of the gradient saturation problem (26). To help the network attend to each scale, previous attention-gated classifier networks have used a staged learning routine (23,28,29). Staged training in a CycleGAN is challenging because the generator and discriminator need to progress in unison. VAE is used as a CycleGAN-compatible alternative to staged learning (24,30,31). VAE facilitates convergence of the core network and attention mechanism while simultaneously allowing the discriminator network to progress in synchrony with the generator network.

The VAE uses an information bottleneck convolutional layer followed by three deconvolutional decoding layers to reproduce the discriminator input (30,31). Figure 2 shows the discriminator network.

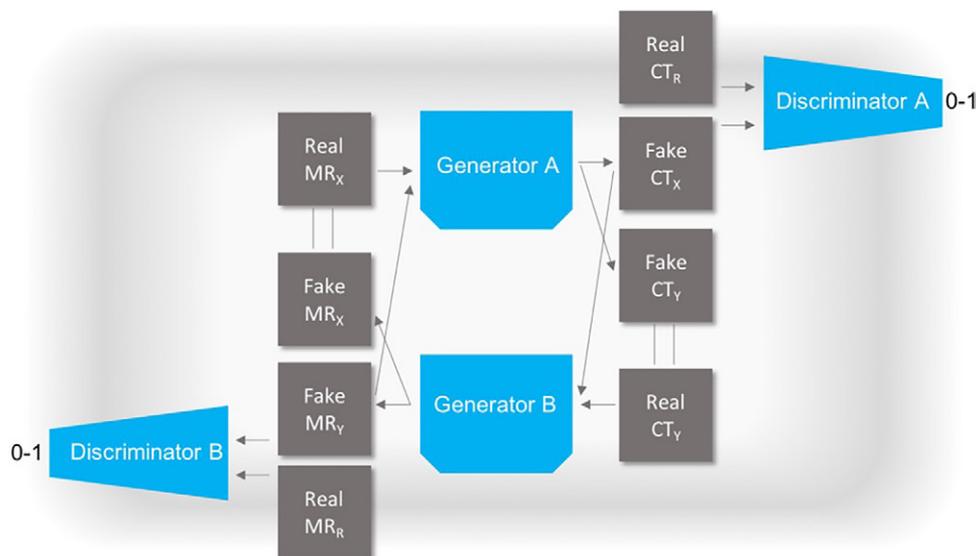
### Training and CycleGAN Framework

This work uses the general CycleGAN framework described by Zhu et al and Jin et al (20,21). Our model is composed of four



**Figure 1:** The gating mechanism used in the discriminator network is shown for the gating signal  $a_2$ , input signal  $a_1$ , and resulting gated signal  $a_g$ . ReLU = rectified linear unit.





**Figure 3:** The overall workflow for the cycle-consistent generative adversarial network architecture with attention-aware discrimination.

#### PSNR, MAE, and SSIM Scores for U-Net, GAN, CycleGAN, and A-CycleGAN Methods

Measure	U-Net	GAN	CycleGAN	A-CycleGAN
<b>PSNR</b>				
Mean	54.67 (95% CI: 53.65, 55.69)	58.81 (95% CI: 58.00, 59.63)	61.58 (95% CI: 61.09, 62.06)	62.35 (95% CI: 61.80, 62.90)
<i>P</i> value	$1.73 \times 10^{-6}$	$3.18 \times 10^{-6}$	.057096	...
<b>MAE</b>				
Mean	26.12 (95% CI: 25.04, 27.19)	23.32 (95% CI: 22.01, 24.63)	21.34 (95% CI: 20.41, 22.27)	19.61 (95% CI: 18.83, 20.39)
<i>P</i> value	$1.92 \times 10^{-6}$	$5.31 \times 10^{-5}$	.010444	...
<b>SSIM</b>				
Mean	0.672 (95% CI: 0.603, 0.741)	0.712 (95% CI: 0.680, 0.744)	0.742 (95% CI: 0.707, 0.777)	0.778 (95% CI: 0.758, 0.798)
<i>P</i> value	.008730	.001593	.049498	...

Note.—The *P* values compare each of the alternative methods to A-CycleGAN using a Wilcoxon signed-rank two-sided test with a sample size of 30. A-CycleGAN = attention-aware CycleGAN, CI = confidence interval, CycleGAN = cycle-consistent GAN, GAN = generative adversarial network, MAE = mean absolute error, PSNR = peak signal-to-noise ratio, SSIM = structure similarity index metric.

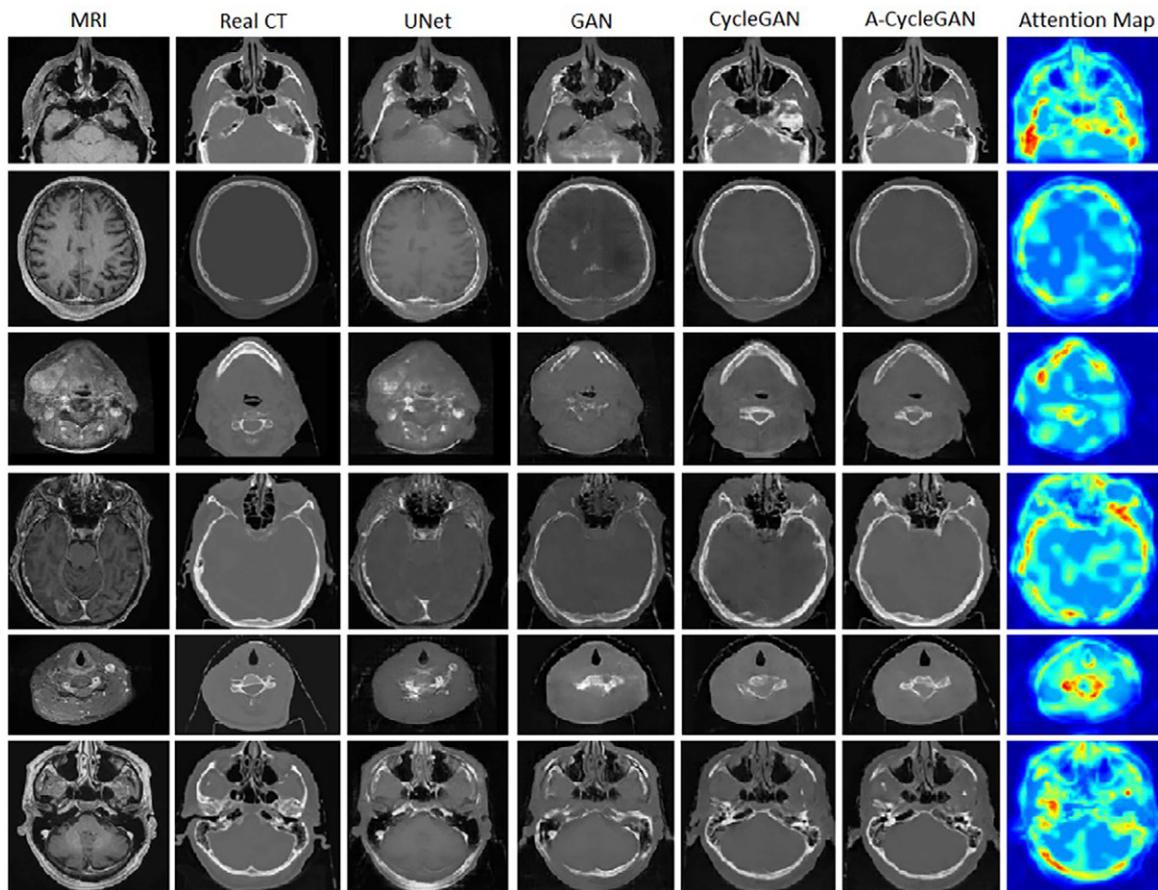
(SSIM) (33). SSIM is an image-quality assessment that is based on the degradation of structural, luminance, and contrast information between two images.

The Table shows the PSNR, MAE, and SSIM scores for the U-Net, GAN, CycleGAN, and A-CycleGAN methods. From the Anderson-Darling test, we determined that a normal distribution could not be assumed for our data. A Wilcoxon signed-rank two-sided test was used to compute comparator *P* values between A-CycleGAN and each alternative method. *P* values less than .05 were considered to be significant.

The Table indicates that the U-Net method had the worst performance for all evaluation metrics. Although the GAN method had some improvement over the U-Net method, it still performed poorly on PSNR, SSIM, and MAE. Although the

GAN method uses adversarial loss, it still relies on some conventional generator loss, so it is sensitive to anatomic misalignment. The CycleGAN method performed much better than the U-Net and GAN methods. A-CycleGAN had the best performance for all evaluation metrics.

Figure 4 shows the visual differences among the real MR, real CT, and predicted CT slices for each of the algorithms. Although the U-Net and GAN methods performed well in some regions of the images, they tended to produce image artifacts or blurred anatomy in some circumstances. The CycleGAN method produced realistic-looking images but still introduced some imaging artifacts. The A-CycleGAN method produced the most realistic CT slices compared with the alternative methods. The attention maps shown in Figure



**Figure 4:** The real MR images, corresponding real CT images, and fake CT images are shown for all methods (U-Net, generative adversarial network [GAN], cycle-consistent GAN [CycleGAN], and attention-aware CycleGAN [A-CycleGAN]) for patients 1 (top) through 6 (bottom) of the holdout test set. The composite attention maps are shown for A-CycleGAN algorithm for all six patients.

4 represent aggregate attention information from all multiscale levels. The attention mechanism works in unison with the primary discriminator pipeline and selectively highlights information propagation based on the interplay between all multiscale levels.

Figure 5 shows the adversarial training loss for discriminators A and B, illustrating how the two opposing network portions progress in unison. The discriminator loss quickly plateaus because the discriminators and generators compete. The loss function is relatively linear compared with conventional CNN loss because the generators get progressively better at creating fake images, whereas the discriminators get progressively better at distinguishing fake images.

## Discussion

This study details the model architecture and learning routine associated with VAE-enhanced A-CycleGANs for MR-to-CT image translation. A-CycleGANs have been used to solve non-medical image-to-image translation, but this the first, to our knowledge, implementation of A-CycleGANs to this problem space. Additionally, to our knowledge, this is the first implementation of VAE-enhanced attention-aware discrimination in any problem space, including nonmedical applications.

From the results, it is clear that unpaired CycleGANs are a superior alternative to paired image-based MR-to-CT image

translation. It is also evident that the A-CycleGAN method achieved statistically significant improvement over alternative methods for MAE and SSIM.

Although direct comparison with the results of previous studies is not possible, previous studies that use CycleGANs report similar improvement over conventional GANs and U-Net–based algorithms when image alignment is difficult or impossible to achieve (21,34). Furthermore, we compared our algorithm with an adaptation of the original U-Net, conditional GAN, and CycleGAN for consistency (20,35,36).

Our U-Net method generated synthetic CT images by incorporating a six-multiscale-level-deep encoder-decoder architecture with skip connections conjoining the convolutional downsampling and deconvolutional upsampling stages. The U-Net model used a dropout layer of 0.35, batch normalization, and ReLU activation at every convolutional layer and did not use a discriminator network. Our GAN architecture used a five-multiscale-level-deep U-Net model as the generator and a three-multiscale-level-deep PatchGAN encoding network as the discriminator. The CycleGAN model used a pair of five-multiscale-level-deep U-Net architectures as the generators and a pair of three-multiscale-level-deep PatchGAN encoding networks for the discriminators. All networks used tanh activation on the last layer of their generators, and all discriminator networks derived their adversarial loss from the raw activations of their last layers.

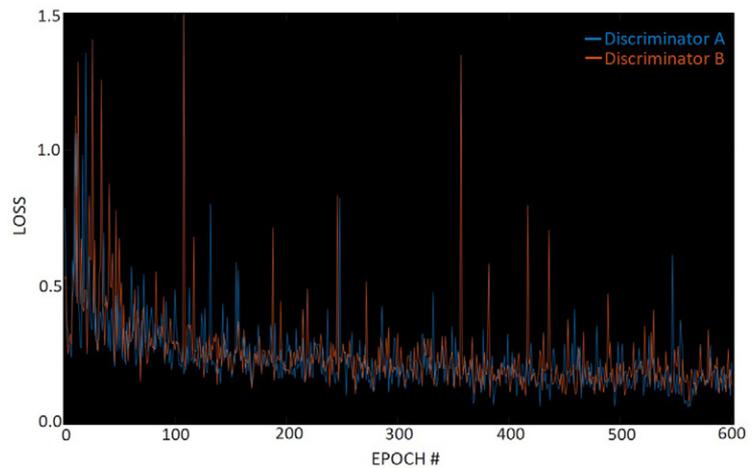
The GAN and CycleGAN models used synchronized batch normalization, leaky ReLU, and no dropout.

GANs excel at problems that do not have a well-defined analytical evaluation metric. If the quality of a prediction cannot be completely captured by a standard loss function, then GANs might be a suitable alternative. GANs encourage realistic predictions and are well suited for medical problems that rely on qualitative clinical decisions that cannot be reduced to combinations of simple analytics such as mean squared error, intersection, union, or edge distances.

Nominally, adversarial discrimination will asymptotically approach 50% accuracy for fake images. Under this ideal circumstance, the network cannot distinguish between real and fake CT images because the fake CT images are ideal translations of their MR counterparts. However, GAN discriminators are usually very simple in order to facilitate model convergence in the context of adversarial training. In contrast, top-performing classification CNN architectures can exceed 200 layers, which dwarfs current CycleGAN discriminators, which consist of five convolutional layers. A network consisting of five convolutional layers will converge quickly but may not have the same predictive power as current state-of-the-art stand-alone classification CNN architectures. VAEs facilitate model convergence and enable deeper architectures without compromising adversarial training. Furthermore, the attention mechanism allows for parsimonious use of network parameters by focusing the network on relevant foreground and background image regions.

To be most effective, this technique might require a change in the clinical workflow, as many patients have undergone CT prior to undergoing MRI. However, many patients undergo MRI prior to undergoing a CT planning scan because MR images help physicians identify the extent of disease and can be used to aid in sensitive tissue delineation. Furthermore, patients will be spared unnecessary radiation doses in cases that enable an MR scan to completely replace a CT scan.

This study had some limitations. The image-to-image translation algorithm uses axial slices that are only  $256 \times 256$ , so the field of view is not big enough to encompass treatment sites that require larger image sizes. Because GANs are sensitive to hyperparameters, it cannot be assumed that our training routine, architecture, and overall approach would generalize well to other larger fields of view. Additionally, this study only considered patients with head, neck, and brain cancer, so we cannot assume this model will generalize well to other body parts. Similarly, we do not know how well this model will transfer to separate but related tasks. MRI is more expensive than CT and is a somewhat limited resource in North America and elsewhere, so this technique may not be practical at every institution. That said, it can certainly be more cost-effective than performing both CT and MRI, which is routinely performed in radiation therapy applications. Although our method achieved a statistically significant improvement compared with alternative methods, the readership would still have to decide if the performance improvement that our algorithm offers is worth the transition from their current solution. In spite of the limitations of this study, unpaired



**Figure 5:** The loss for discriminator A (blue) and discriminator B (orange) is shown for the entire training routine, which consists of 600 epochs.

image-to-image translation solutions could prove useful in many medical imaging applications, as variations in patient anatomy and setup positions are a ubiquitous problem in therapeutic and diagnostic imaging. Additionally, MR-to-CT image translation could have other use cases, such as in perioperative planning, but those use cases would first have to be clinically validated. In addition to unpaired image-to-image translation problems, such as CT-to-cone beam CT image translation and MR-to-CT image translation, CycleGANs allow models to rely on pure adversarial loss while simultaneously making inferences based on their input image conditions (37,38).

In summary, this study demonstrated that unpaired A-CycleGANs enhanced with VAE are a superior alternative to current state-of-the-art MR-to-CT image translation methods.

**Author contributions:** Guarantors of integrity of entire study, V.K., T.W., O.M., S.S.Y., T.D.S.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, V.K., A.P., T.W., O.M., T.D.S.; clinical studies, B.P.Z., O.M., S.S.Y.; experimental studies, V.K., A.P., T.W., S.S.Y., T.D.S.; statistical analysis, V.K., T.W., L.M., O.M., T.D.S.; and manuscript editing, V.K., B.P.Z., L.M., O.M., S.S.Y., T.D.S.

**Disclosures of Conflicts of Interest:** V.K. Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: employed by 1983. Other relationships: disclosed no relevant relationships. B.P.Z. disclosed no relevant relationships. A.P. disclosed no relevant relationships. T.W. disclosed no relevant relationships. J.W.C. disclosed no relevant relationships. L.M. disclosed no relevant relationships. O.M. disclosed no relevant relationships. S.S.Y. disclosed no relevant relationships. T.D.S. disclosed no relevant relationships.

## References

1. Webb S. Intensity-modulated radiation therapy. Boca Raton, Fla: CRC, 2015.
2. Jelen U, Alber M. A finite size pencil beam algorithm for IMRT dose optimization: density corrections. *Phys Med Biol* 2007;52(3):617–633.
3. De Meerleer G, Villeirs G, Bral S, et al. The magnetic resonance detected intraprostatic lesion in prostate cancer: planning and delivery of intensity-modulated radiotherapy. *Radiother Oncol* 2005;75(3):325–333.
4. Kearney V, Descovich M, Sudhyadhom A, Cheung JP, McGuinness C, Solberg TD. A continuous arc delivery optimization algorithm for CyberKnife m6. *Med Phys* 2018;45(8):3861–3870.
5. Kearney V, Solberg T, Jensen S, Cheung J, Chuang C, Valdes G. Correcting TG 119 confidence limits. *Med Phys* 2018;45(3):1001–1008.

6. Kearney V, Cheung JP, McGuinness C, Solberg TD. CyberArc: a non-coplanar-arc optimization algorithm for CyberKnife. *Phys Med Biol* 2017;62(14):5777–5789.
7. Kearney V, Huang Y, Mao W, Yuan B, Tang L. Canny edge-based deformable image registration. *Phys Med Biol* 2017;62(3):966–985.
8. Kearney V, Chen S, Gu X, et al. Automated landmark-guided deformable image registration. *Phys Med Biol* 2015;60(1):101–116.
9. Interian Y, Rideout V, Kearney VP, et al. Deep nets vs expert designed features in medical physics: An IMRT QA case study. *Med Phys* 2018;45(6):2672–2680.
10. Han X. MR-based synthetic CT generation using a deep convolutional neural network method. *Med Phys* 2017;44(4):1408–1419.
11. Nie D, Cao X, Gao Y, Wang L, Shen D. Estimating CT image from MRI data using 3D fully convolutional networks. *Deep Learn Data Label Med Appl* (2016) 2016;2016:170–178.
12. Kearney V, Chan JW, Haaf S, Descovich M, Solberg TD. DoseNet: a volumetric dose prediction algorithm using 3D fully-convolutional neural networks. *Phys Med Biol* 2018;63(23):235022.
13. Kearney V, Chan J, Descovich M, Yom S, Solberg T. A multi-task CNN model for autosegmentation of prostate patients. *Int J Radiat Oncol Biol Phys* 2018;102(3):S214.
14. Kearney V, Chan JW, Valdes G, Solberg TD, Yom SS. The application of artificial intelligence in the IMRT planning process for head and neck cancer. *Oral Oncol* 2018;87:111–116.
15. Kearney V, Chan JW, Wang T, Perry A, Yom SS, Solberg TD. Attention-enabled 3D boosted convolutional neural networks for semantic CT segmentation using deep supervision. *Phys Med Biol* 2019;64(13):135001.
16. Chan JW, Kearney V, Haaf S, et al. A convolutional neural network algorithm for automatic segmentation of head and neck organs at risk using deep lifelong learning. *Med Phys* 2019;46(5):2204–2213.
17. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. Paper presented at: *Advances in Neural Information Processing Systems 27 (NIPS 2014)*; December 8–13, 2014; Montreal, Canada. <https://papers.nips.cc/paper/5423-generative-adversarial-nets>.
18. Nie D, Trullo R, Lian J, et al. Medical image synthesis with context-aware generative adversarial networks. In: Descoteaux M, Maier-Hein L, Franz A, Jannin P, Collins D, Duchesne S, eds. *Medical image computing and computer assisted intervention – MICCAI 2017*. Vol 10435, Lecture Notes in Computer Science. Cham, Switzerland: Springer, 2017; 417–425.
19. Kearney V, Chan J, Haaf S, Yom S, Solberg T. PO-0997 a synthetic generative adversarial network for semantic lung tumor segmentation. *Radiother Oncol* 2019;133(suppl 1):S549.
20. Zhu JY, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. ArXiv 1703.10593 [preprint] <https://arxiv.org/abs/1703.10593>. Posted March 30, 2017. Accessed February 2019.
21. Jin CB, Jung W, Joo S, et al. Deep CT to MR synthesis using paired and unpaired data. ArXiv 1805.10790 [preprint] <https://arxiv.org/abs/1805.10790>. Posted May 28, 2018. Accessed February 2019.
22. Mejjati YA, Richardt C, Tompkin J, Cosker D, Kim KI. Unsupervised attention-guided image to image translation. ArXiv 1806.02311 [preprint] <https://arxiv.org/abs/1806.02311>. Posted June 6, 2018. Accessed February 2019.
23. Schlemper J, Oktay O, Schaap M, et al. Attention gated networks: learning to leverage salient regions in medical images. ArXiv 1808.08114 [preprint] <https://arxiv.org/abs/1808.08114>. Posted August 22, 2018. Accessed February 2019.
24. Nash C, Williams CKI. The shape variational autoencoder: a deep generative model of part-segmented 3D objects. *Comput Graph Forum* 2017;36(5):1–7.
25. Kastaniotis D, Ntinou I, Tsourounis D, Economou G, Fotopoulos S. Attention-aware generative adversarial networks (ATA-GANs). Paper presented at: 2018 IEEE 13th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP); June 10–12, 2018; Piscataway, NJ.
26. Oktay O, Schlemper J, Folgoc LL, et al. Attention U-Net: learning where to look for the pancreas. ArXiv:1804.03999 [preprint] <https://arxiv.org/abs/1804.03999>. Posted April 11, 2018. Accessed February 2019.
27. Luong MT, Pham H, Manning CD. Effective approaches to attention-based neural machine translation. ArXiv:1508.04025 [preprint] <https://arxiv.org/abs/1508.04025>. Posted August 17, 2015. Accessed February 2019.
28. Jetley S, Lord NA, Lee N, Torr PH. Learn to pay attention. ArXiv 1804.02391 [preprint] <https://arxiv.org/abs/1804.02391>. Posted April 6, 2018. Accessed February 2019.
29. Ypsilantis PP, Montana G. Learning what to look in chest x-rays with a recurrent visual attention model. ArXiv 1701.06452 [preprint] <https://arxiv.org/abs/1701.06452>. Posted January 23, 2017. Accessed February 2019.
30. Doersch C. Tutorial on variational autoencoders. ArXiv 1606.05908 [preprint] <https://arxiv.org/abs/1606.05908>. Posted June 19, 2016. Accessed February 2019.
31. Kearney V, Haaf S, Sudhyadhom A, Valdes G, Solberg TD. An unsupervised convolutional neural network-based algorithm for deformable image registration. *Phys Med Biol* 2018;63(18):185017.
32. Wu Y, He K. Group normalization. ArXiv 1803.08494 [preprint] <https://arxiv.org/abs/1803.08494>. Posted March 22, 2018. Accessed February 2019.
33. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process* 2004;13(4):600–612.
34. Welander P, Karlsson S, Eklund A. Generative adversarial networks for image-to-image translation on multi-contrast MR images - a comparison of CycleGAN and UNIT. ArXiv 1806.07777 [preprint] <https://arxiv.org/abs/1806.07777>. Posted June 20, 2018. Accessed February 2019.
35. Isola P, Zhu JY, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. ArXiv 1611.07004 [preprint] <https://arxiv.org/abs/1611.07004>. Posted November 21, 2016. Accessed February 2019.
36. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells W, Frangi A, eds. *Medical image computing and computer-assisted intervention – MICCAI 2015*. Vol 9351, Lecture Notes in Computer Science. Cham, Switzerland: Springer, 2015; 234–241.
37. Kida S, Kaji S, Nawa K, et al. Cone-beam CT to planning CT synthesis using generative adversarial networks. ArXiv 1901.05773v1 [preprint] <https://arxiv.org/abs/1901.05773v1>. Posted January 17, 2019. Accessed February 2019.
38. Januszewski M, Jain V. Segmentation-enhanced CycleGAN. *Biorxiv* 10.1101/548081v1 [preprint] <https://www.biorxiv.org/content/10.1101/548081v1>. Posted February 13, 2019. Accessed February 2019.