MULTI-DOMAIN UNPAIRED ULTRASOUND IMAGE ARTIFACT REMOVAL USING A SINGLE CONVOLUTIONAL NEURAL NETWORK

Jaeyoung Huh, Shujaat Khan, and Jong Chul Ye, Fellow, IEEE

Korea Advanced Institute of Science and Technology (KAIST)
Dept of Bio and Brain Engineering
Daejeon 34141, Republic of Korea

ABSTRACT

Ultrasound imaging (US) often suffers from distinct image artifacts from various sources. Classic approaches for solving these problems are usually model-based iterative approaches that have been developed specifically for each type of artifact, which are often computationally intensive. Recently, deep learning approaches have been proposed as computationally efficient and high performance alternatives. Unfortunately, in the current deep learning approaches, a dedicated neural network should be trained with matched training data for each artifact type. This poses a fundamental limitation in the practical use of deep learning for US, since large number of paired data is required for supervised training of multiple models to deal with various US image artifacts. Inspired by the recent success of multi-domain image transfer, herein, we propose a novel unpaired deep learning approach where a single neural network can deal with different types of US artifacts simply by changing a mask vector that switches between different target domains. The proposed method can generate high quality images by removing distinct artifacts, which are comparable to those obtained by separately trained multiple neural networks.

***Index Terms*—** Ultrasound Imaging, Deep learning, Multi-domain translation, Deconvolution, Speckle removal

1. INTRODUCTION

In contrast to computed tomography (CT) and magnetic resonance imaging (MRI), ultrasound imaging (US) poses no radiation risks to the patient and enjoys fast acquisition time, while the hardware system is much simpler. Therefore, US is very useful for many clinical and portable diagnostic applications. Unfortunately, US suffers from imaging artifact such as image blur, aberration etc. Moreover, when there are small structures, the speckle patterns generated from the signal interference sometimes hinder the accurate diagnosis. For the last few decades, many researchers have proposed various model-based iterative algorithms to address these problems [1–4]. While the results are impressive, these model-based approaches are computationally expensive. To deal with the challenges of conventional model-based approaches, a variety of machine learning approaches have been proposed recently. For example, in [5], deep learning-based beamformers have been suggested as promising alternatives to the delay-and-sum (DAS) or adaptive beamformers. In [6, 7], speckle denoising beamformers using convolutional neural network were suggested. Similarly, various US artifact removal algorithms have been also implemented using deep neural networks [8–10].

Despite these promising efforts, the deep neural network approaches for US imaging still have some technical hurdles for their wide acceptance. US images are usually corrupted with different types of artifacts, and each user often prefers distinct choice of artifact suppression algorithms depending on the clinical applications. For example, to maximize the segmentation performance of cardiac walls, lesions, etc, the speckle noises should be reduced, on the other hand speckle statistics have important information for diagnosis the functional characteristic [11, 12]. This implies that a variety of neural networks should be stored in an US scanner to deal with various image artifacts to satisfy the users' demands. Furthermore, conventional approaches can only utilize one type of data for a specific model training. This limitation hinders the learning from different targets which can provide useful information for better generalization.

To address the aforementioned problems, herein, we propose a novel deep learning approach to overcome these fundamental limitations. In our method, a *single* neural network can be trained to deal with different types of image artifacts by simply changing the target mask vector. Since multiple US artifact reduction problems can be solved by a single neural network, this is expected to reduce the burden caused by multiple model designing and by exploiting different target's data, better generalization can be achieved.

2. PROPOSED METHOD

2.1. Background

The proposed method is inspired by the recently proposed multi-domain image transfer models such as StarGAN and CollaGAN [13, 14]. Although these models use a single generator to translate an image to multiple domains, they often require back and forth (bi-directional) translation which

This work was supported by the National Research Foundation of Korea under Grant NRF-2020R1A2B5B03001980.

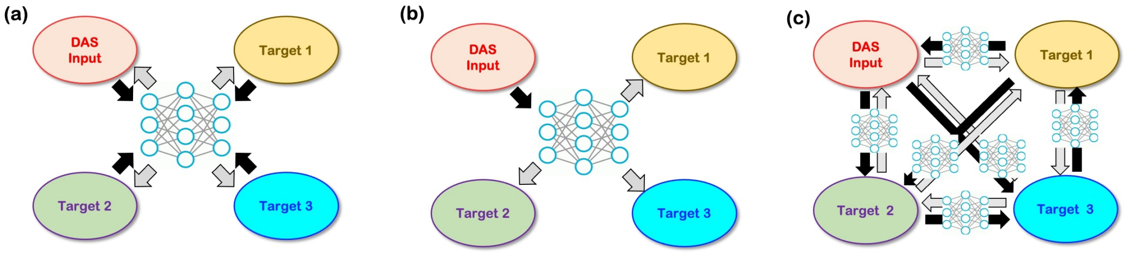


Fig. 1. Multi-domain image translation at the test phase using (a) StarGAN, (b) our method, and (c) CycleGAN. For the case StarGAN, a single generator should translate between any domains in forward and backward directions, whereas in our method, a trained generator translates only from DAS to any domains.

is unnecessary as the US image artifacts removal problem is a uni-directional translation problem and it can be solved with relatively less complexity. More specifically, while StarGAN should translate each domain to every other domain as shown in Fig. 1(a), this is not necessary in US artifact removal problems, since DAS images are raw data obtained from the scanner and it is not necessary to re-generate DAS images from other domain images. Therefore, a correct image reconstruction pipeline should be asymmetric as shown in Fig. 1(b), where artifact suppressed images are generated only from DAS data. It turns out that the lack of symmetry makes the multi-domain translation task much simpler, thereby significantly reducing the network complexity and improving the performance.

2.2. Dataset

We used *in vivo* and phantom images to train the network. The *in vivo* images are focused B-mode images acquired using a linear probe (L3-12H) from the US system E-CUBE 12R (Alphinion, Korea). The images were scanned from four parts of the carotid and thyroid areas of 10 volunteers. The *in vivo* has 10 frames for each body part. The phantom images are tissue mimicking phantoms that are also acquired using focused imaging mode using a linear array probe. Both the *in vivo* and phantom images are taken with a center frequency of 8.5 MHz. For the training dataset, we selected 125 phantom images and 8 volunteers' data which comprises of 304 *in vivo* images. For target output, high quality tissue-reflectively function (deconvoluted images) and speckle-free filtered images are generated from original DAS images using sparse-reconstruction model [1] and non-local low-rank-based speckle denoising model [3], respectively.

2.3. Loss function

One key difference compared to the CycleGAN geometry is that the target distribution is modeled as the mixture distribution of multiple target distributions, which can be separated by a classifier K_η that maps each sample in $\mathcal{X}$ to discrete domain labels. Thus, our goal is to find a generator $G_\theta : \mathcal{Y} \mapsto \mathcal{X}$ which enables the correct classification using K_η. Here, $\mathcal{Y}$ is original DAS image domain and $\mathcal{X}$ is multiple target domain.

Our multi-domain artifact removal problem can be formulated by

$$\min_{\theta,\phi,\eta} \max_{\varphi,\psi} \ell_{mlt}(\theta,\phi,\eta;\varphi,\psi), \tag{1}$$

where

$$\begin{aligned}\ell_{mlt}(\theta,\phi,\eta;\varphi,\psi) :=& \lambda_{cyc}\ell_{cycle}(\theta,\phi) + \ell_{Disc}(\theta,\phi;\varphi,\psi) \\ &+ \lambda_{GP}\ell_{GP}(\varphi,\psi) + \lambda_{cls}\ell_{cls}(\theta,\eta).\end{aligned} \tag{2}$$

where each λ is the weighting parameter depending on the relative importance. We used Wasserstein GAN with Gradient Penalty (WGAN-GP) [15, 16] for optimization formulation.

In (2), the cycle-consistency term is given by

$$\ell_{cycle} = E_{x\sim P_x}\|x - G_\theta(F_\phi(x))\| + E_{y\sim P_y}\|y - F_\phi(G_\theta(y))\|,$$

whereas the discriminator term is

$$\begin{aligned}\ell_{Disc} =& E_{x\sim P_x}[D_\varphi(x)] - E_{y\sim P_y}[D_\varphi(G_\theta(y))] \\ &+ E_{y\sim P_y}[D_\psi(y)] - E_{x\sim P_x}[D_\psi(F_\phi(x))].\end{aligned}$$

The gradient penalty term is

$$\begin{aligned}\ell_{GP} =& - E_{x\sim P_x}[(\|\nabla_{\tilde{x}} D_\varphi(\tilde{x})\|_2 - 1)^2] \\ &- E_{y\sim P_y}[(\|\nabla_{\tilde{y}} D_\psi(\tilde{y})\|_2 - 1)^2],\end{aligned}$$

where $\tilde{x} = \alpha x + (1-\alpha)G_\theta(y)$ and $\tilde{y} = \alpha y + (1-\alpha)F_\phi(x)$ with α being the random variables from the uniform distribution between $[0, 1]$ [16].

Lastly, the classification loss term is defined by

$$\begin{aligned}\ell_{cls} =& - E_{x\sim P_x}[p(x)\log K_\eta(x)] \\ &- E_{y\sim P_y}[p(G_\theta(y))\log K_\eta(G_\theta(y))].\end{aligned}$$

where the output of the classifier K_η is the probability distribution indicating the probability belonging to the target domains, and $p(x)$ is the one-hot vector encoded true label probability for x. The $p(G_\theta(y))$ is one-hot vector encoded target label for the generator G_θ for a given DAS image y.

2.4. Implementation Details

For model training, the hyper-parameters were set to be $\lambda_{cyc} = 20$, $\lambda_{GP} = 30$ and $\lambda_{cls} = 1$ which shows best results.

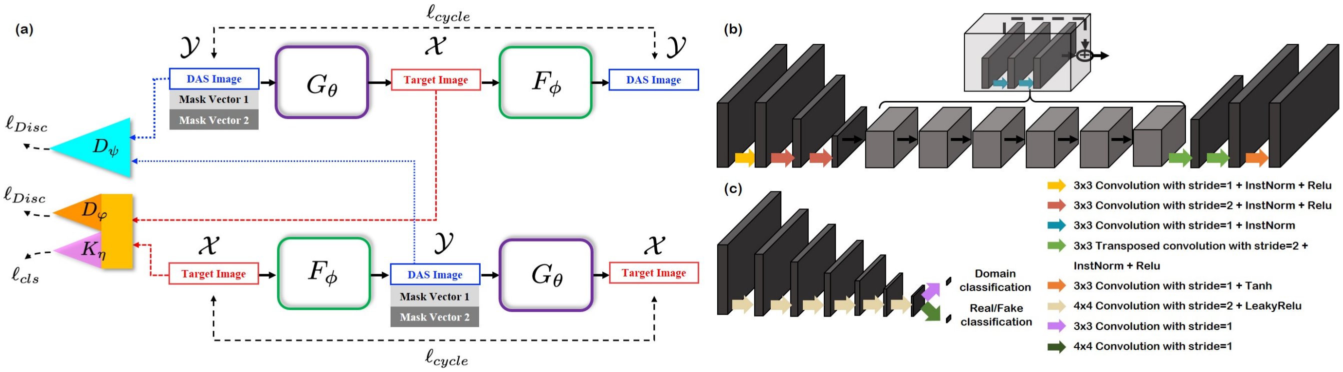


Fig. 2. Proposed multi-domain unpaired artifact removal networks.

The model was trained for 1000 epochs using the Adam optimizer with the batch size of 4 samples and initial learning-rate of $1e-4$ which linearly decreased after 500 epochs. To avoid overfitting, augmentation technique like flipping, rotating and random scaling were also incorporated. The neural network models were trained using Python 3.8 with Tensorflow 1.13.1 on an NVIDIA GeForce GTX 1080 Ti GPU about 21 hours, while classic deconvolution and despeckling models were implemented using MATLAB 2017a on a Intel i7 processor and 64 GB of memory.

2.5. Neural Network Implementation

Fig. 2 illustrates the proposed multi-domain unpaired artifact removal network originated from the optimization problem in (2). The network architecture consists of two generators. The main generator G_θ converts DAS images to artifact-free target domain images. The second generator F_ϕ returns the target domain images back to the original DAS domain images. Once the model is fully trained, we only use G_θ at the inference phase.

Since the primary idea is to indicate the target domain using an input mask vector, m:

$$x = G_\theta(y; m) \tag{3}$$

Here, care should be taken due to the existence of the classifier e.g, a DAS input image with the mask vector $m = [1, 0]$ creates the deconvolution image, while $m = [0, 1]$ will generate despeckled image. The mask vector is concatenated along the channel direction with the input image. On the other hand, F_ϕ does not need a mask vector since there exists no classifier in $\mathcal{Y}$.

Additionally, there are two discriminator networks D_φ and D_ψ as shown in Fig. 2. Specifically, the network D_φ tries to find the difference between the true image x and the generated image $G_\Theta(y)$, whereas the network D_ψ attempts to find the fake measurement data that are generated by the synthetic measurement procedure $F_\phi(x)$.

Finally, we have the domain classifier K_η which distinguishes between deconvolution and despeckled images. In fact, the domain classifier and discriminators are both classifiers, so their structure share many commonalities. Therefore, as shown in Fig. 2(b), the discriminator D_φ and the classifier K_η are implemented using a same network architecture with double output headers composed of a domain classifier or discriminator.

3. EXPERIMENTAL RESULTS

All models were evaluated for reconstruction performances using peak-signal to noise ratio (PSNR), structural similarity (SSIM), contrast ratio (CR), contrast-to-noise-ratio (CNR) and generalized contrast-to-noise ratio (GCNR) metrics [17].

3.1. Baseline Algorithms

For comparative study, the following algorithms were used as baselines: 1) supervised learning, 2) CycleGAN [18], and 3) StarGAN [13]. All three algorithms used the same generator architecture as the proposed method. However, there is no domain classification loss in CycleGAN. Therefore, the discriminator used in CycleGAN has different architecture in the last layer, i.e. with a single head instead of a multi-head.

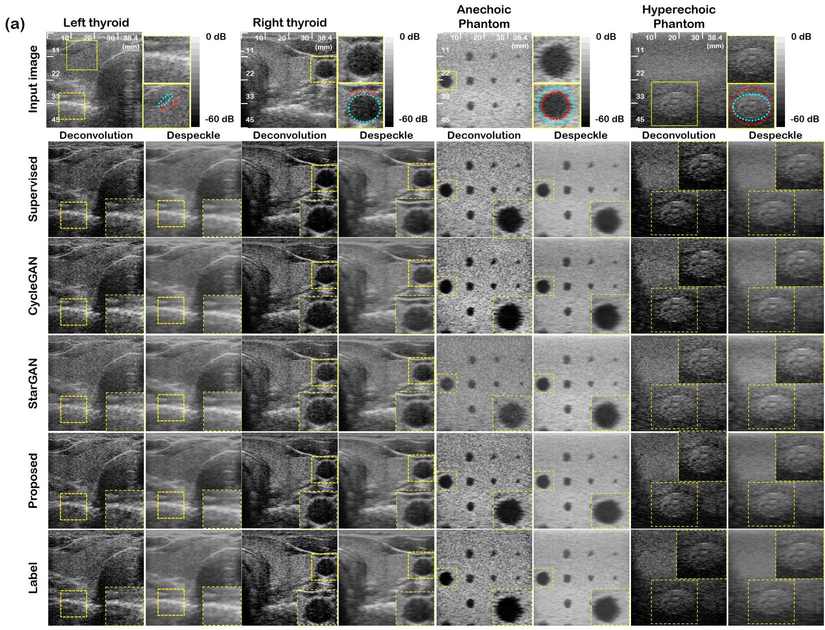
Table 1. Algorithm Configuration

Details	Supervised	CycleGAN	StarGAN
Dataset	Paired	Unpaired	Unpaired
Normalization	Intensity of the image as - 1 to 1		
Data Augmentation	Flipping, rotating and random scaling		
Optimizer	Adam optimizer		
Learning Rate	Linearly decreasing from 1e-4 after the half of total epoch.		
Batch Size	4		
Epoch	200	1000	
Parameters($\lambda_{GP}, \lambda_{cls}, \lambda_{rec}$)	None	(30,None,0.5)	(30,1,20)
Others	None	WGAN-GP	

3.2. Algorithm comparison

The test dataset are composed of *in vivo* and phantom images. The *in vivo* dataset consists of 80 frames of images from four different body parts of two volunteers. The phantom dataset of 16 images consists of 10 anechoic and 6 hyperechoic phantom images.

Fig. 3 shows the comparison results using various algorithms for the conversion to the deconvolution and despeckled image domain. All the deconvolution output images show improved contrast. While the StarGAN output has slightly improved its contrast compared to the input image, the output



(b)

Deconvolution performance evaluation

	In vivo					Phantom				
	PSNR	SSIM	CNR	GCNR	CR	PSNR	SSIM	CNR	GCNR	CR
a	17.15	0.68	**0.82**	0.52	7.97	15.27	0.74	1.25	**0.60**	11.41
b	X	X	0.74	0.53	**9.15**	X	X	1.13	0.56	12.24
c	**24.35**	0.79	0.79	**0.55**	9.13	21.48	0.82	**1.26**	**0.60**	**14.01**
d	23.97	**0.80**	0.75	0.53	8.61	21.54	0.81	1.14	0.56	13.84
e	20.07	0.72	0.77	0.50	7.53	19.07	0.76	1.15	0.57	9.59
f	23.86	0.79	0.76	0.53	8.98	**22.13**	**0.84**	1.19	0.58	13.03

(c)

Despeckle performance evaluation

	In vivo					Phantom				
	PSNR	SSIM	CNR	GCNR	CR	PSNR	SSIM	CNR	GCNR	CR
a	26.99	0.65	0.82	0.52	**7.97**	26.68	0.66	1.25	0.60	11.41
b	X	X	1.02	**0.62**	7.52	X	X	1.46	0.65	10.79
c	**35.58**	**0.93**	1.04	**0.62**	7.29	29.68	**0.90**	**1.49**	**0.67**	11.98
d	29.62	0.89	**1.05**	0.61	7.38	19.96	0.84	1.39	0.64	11.97
e	31.73	0.89	0.99	0.59	7.57	28.74	0.87	1.38	0.65	11.34
f	31.78	0.91	1.02	0.61	7.86	**30.02**	0.89	**1.49**	**0.67**	**12.52**

(a : Input, b : Label, c : Supervised, d : CycleGAN, e : StarGAN, f : Proposed)

PSNR : Peak Signal to Noise Ratio / SSIM : Structural Similarity

CNR : Contrast Noise Ratio / GCNR : Generalized Contrast Noise Ratio / CR : Contrast Ratio

Fig. 3. (a) Qualitative comparison results with two *in vivo* and two phantom dataset using various algorithms. The yellow box under each image shows the magnified image of yellow dot-line box. The red and blue lines are the selected region for calculating contrast metrics. (b) Deconvolution quantitative comparison with test set. (c) Despeckle quantitative comparison with test set. The results highlighted in red denote the best score and the blue denote the second best.

from the proposed method is closer to the label image. Moreover, the side-by-side comparison with supervised and CycleGAN approach show that the proposed method provides qualitatively comparable results. However, it is remarkable that the deconvolution and despeckled images using supervised learning and CycleGAN are generated from independent network trained separately, whereas the proposed method generated both images with a single generator.

The despeckle case also shows well-suppressed speckle image. After removing the speckle pattern, the boundaries of structure become smooth in comparative methods. However, as shown in the Fig. 3, the results of the proposed method preserved the structure boundaries better than the StarGAN.

In the Fig. 3, each table shows the results of individual algorithms. The input is an original DAS image and the labels, generated by the classical algorithms, are assumed to be the target/ground-truth images. We show the result for *in vivo* dataset and phantom datasets separately. We did not calculate the PSNR, SSIM for the label images, because the label images were used as reference for the output images.

The deconvolution process not only increases the resolution but it also enhances the noise component. The CNR and GCNR values for deconvolution tasks are slightly decreased compared to DAS input. On the other hand, the CR values, which are insensitive to noise boosting, are noticeably increased. This drop in CNR and GCNR metrics is only due to the choice of a particular deconvolution method and it can be easily resolved by choosing a better target method.

3.3. Computational Time

Another important metric for performance is the computation time. Herein, a CPU run-time is measure by using only CPU implementations of all the algorithms on a computer with an Intel i7 processor and 64 GB memory. On average, to process a single image of size 256 × 256, the proposed method takes only 0.57 seconds irrespective to the type of targeted task, whereas the classical deconvolution model [1], NLLR despeckle [3] took 255.71 and 251.27 seconds, respectively. As the architecture of generator network for supervised learning, CycleGAN, StarGAN methods are basically same as ours, therefore their computational times are also the same as ours.

4. CONCLUSION

Due to the transducer limitations or wave interference, the ultrasound image has many limitations for accurate diagnosis. To overcome this problem, we proposed a multi-domain image artifact removal method. Unlike the CycleGAN architecture, where multiple generator and discriminator networks are required, the proposed method provides a single generator that can be used for multiple artifact removal tasks by simply changing the mask vector. As the proposed method can process images instantaneously without increasing the number of models, it can exploit redundant information from different targets to achieve better generalization. With the noticeably improved results on real *in vivo* and phantom scans, we believe that our method can be used for practical US systems.

5. REFERENCES

[1] Junbo Duan, Hui Zhong, Bowen Jing, Siyuan Zhang, and Mingxi Wan, "Increasing axial resolution of ultrasonic imaging with a joint sparse representation model," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 63, no. 12, pp. 2045–2056, 2016.

[2] Radovan Jirik and Torfinn Taxt, "Two-dimensional blind Bayesian deconvolution of medical ultrasound images," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 55, no. 10, pp. 2140–2153, 2008.

[3] L. Zhu, C. Fu, M. S. Brown, and P. Heng, "A non-local low-rank framework for ultrasound speckle reduction," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017, pp. 493–501.

[4] Pierrick Coupé, Pierre Hellier, Charles Kervrann, and Christian Barillot, "Nonlocal means-based speckle filtering for ultrasound images," *IEEE transactions on image processing*, vol. 18, no. 10, pp. 2221–2229, 2009.

[5] Shujaat Khan, Jaeyoung Huh, and Jong Chul Ye, "Adaptive and compressive beamforming using deep learning for medical ultrasound," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 2020.

[6] Dongwoon Hyun, Leandra L Brickson, Kevin T Looby, and Jeremy J Dahl, "Beamforming and speckle reduction using neural networks," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 66, no. 5, pp. 898–910, 2019.

[7] Shujaat Khan, Jaeyoung Huh, and Jong Chul Ye, "Switchable deep beamformer for ultrasound imaging using adain," in *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2021, pp. 677–680.

[8] Yeo Hun Yoon, Shujaat Khan, Jaeyoung Huh, and Jong Chul Ye, "Efficient B-mode ultrasound image reconstruction from sub-sampled RF data using deep learning," *IEEE transactions on medical imaging*, vol. 38, no. 2, pp. 325–336, 2019.

[9] Priyanka Kokil and S Sudharson, "Despeckling of clinical ultrasound images using deep residual learning," *Computer Methods and Programs in Biomedicine*, p. 105477, 2020.

[10] Shujaat Khan, Jaeyoung Huh, and Jong Chul Ye, "Variational formulation of unsupervised deep learning for ultrasound image artifact removal," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 68, no. 6, pp. 2086–2100, 2021.

[11] Woo Kyung Moon, Chung-Ming Lo, Chiun-Sheng Huang, Jeon-Hor Chen, and Ruey-Feng Chang, "Computer-aided diagnosis based on speckle patterns in ultrasound images," *Ultrasound in medicine & biology*, vol. 38, no. 7, pp. 1251–1261, 2012.

[12] Sergio Mondillo, Maurizio Galderisi, Donato Mele, Matteo Cameli, Vincenzo Schiano Lomoriello, Valerio Zacà, Piercarlo Ballo, Antonello D'Andrea, Denisa Muraru, Mariangela Losi, et al., "Speckle-tracking echocardiography: a new technique for assessing myocardial function," *Journal of Ultrasound in Medicine*, vol. 30, no. 1, pp. 71–83, 2011.

[13] Yunjey Choi, Minje Choi, Munyoung Kim, Jung-Woo Ha, Sunghun Kim, and Jaegul Choo, "StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8789–8797.

[14] Dongwook Lee, Junyoung Kim, Won-Jin Moon, and Jong Chul Ye, "CollaGAN: Collaborative GAN for missing image data imputation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2487–2496.

[15] Martin Arjovsky, Soumith Chintala, and Léon Bottou, "Wasserstein gan," *arXiv preprint arXiv:1701.07875*, 2017.

[16] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C Courville, "Improved training of wasserstein gans," in *Advances in neural information processing systems*, 2017, pp. 5767–5777.

[17] Alfonso Rodriguez-Molares, Ole Marius Hoel Rindal, Jan D'hooge, Svein-Erik Måsøy, Andreas Austeng, Muyinatu A Lediju Bell, and Hans Torp, "The generalized contrast-to-noise ratio: a formal definition for lesion detectability," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 67, no. 4, pp. 745–759, 2019.

[18] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.