Cycle-consistent adversarial denoising network for multiphase coronary CT angiography

Eunhee Kang
Bio Imaging and Signal Processing Laboratory, Department of Bio and Brain Engineering, KAIST, Daejeon, Republic of Korea

Hyun Jung Koo, Dong Hyun Yang, and Joon Bum Seo
Department of Radiology, University of Ulsan College of Medicine, Seoul, Republic of Korea

Jong Chul Ye
Bio Imaging and Signal Processing Laboratory, Department of Bio and Brain Engineering, KAIST, Daejeon, Republic of Korea

(Received 3 July 2018; revised 10 October 2018; accepted for publication 23 October 2018; published 26 December 2018)

Purpose: In multiphase coronary CT angiography (CTA), a series of CT images are taken at different levels of radiation dose during the examination. Although this reduces the total radiation dose, the image quality during the low-dose phases is significantly degraded. Recently, deep neural network approaches based on supervised learning technique have demonstrated impressive performance improvement over conventional model-based iterative methods for low-dose CT. However, matched low- and routine-dose CT image pairs are difficult to obtain in multiphase CT. To address this problem, we aim at developing a new deep learning framework.

Method: We propose an unsupervised learning technique that can remove the noise of the CT images in the low-dose phases by learning from the CT images in the routine dose phases. Although a supervised learning approach is not applicable due to the differences in the underlying heart structure in two phases, the images are closely related in two phases, so we propose a cycle-consistent adversarial denoising network to learn the mapping between the low- and high-dose cardiac phases.

Results: Experimental results showed that the proposed method effectively reduces the noise in the low-dose CT image while preserving detailed texture and edge information. Moreover, thanks to the cyclic consistency and identity loss, the proposed network does not create any artificial features that are not present in the input images. Visual grading and quality evaluation also confirm that the proposed method provides significant improvement in diagnostic quality.

Conclusions: The proposed network can learn the image distributions from the routine-dose cardiac phases, which is a big advantage over the existing supervised learning networks that need exactly matched low- and routine-dose CT images. Considering the effectiveness and practicability of the proposed method, we believe that the proposed can be applied for many other CT acquisition protocols. © 2018 American Association of Physicists in Medicine [https://doi.org/10.1002/mp.13284]

Key words: adversarial neural network, cycle consistency, deep learning, low-dose CT, multiphase coronary CT angiography, unsupervised learning

1. INTRODUCTION

X-ray computed tomography (CT) is one of the most widely used imaging modalities for diagnostic purposes. For example, in cardiac coronary CT angiography (CTA), a series of CT images are acquired while examination is conducted with contrast injection, which helps clinicians to identify heart disease. However, it is often difficult to predict in which heart phase the disease area can be seen better, so multiphase acquisition is often necessary. In the case of valve disease, cardiac motion information from multiphase acquisition is essential, and in evaluating cardiac function, myocardial motion should be evaluated, in which case multiphase acquisition is needed.

However, taking all phase images at full dose is not allowed due to the excessive radiation dose. On the other hand, it is risky to get the entire phase at the low dose because none of the cardiac phases may be diagnostically useful. Therefore, in clinical environments, a multiphase tube current-modulated CTA as shown in Fig. 1 is often used to obtain at least one high-dose phase image, the information of which may also be exploited by the radiologist to interpret the low-dose phase images.

Although this tube current modulation can reduce the total radiation dose, it also introduces noise in the projection data of the low-dose phases. This results in CT images with different noise levels and contrast at different cardiac phases (see Fig. 1). Although model-based iterative reconstruction (MBIR) methods have been developed to address this, the MBIR approaches suffer from relatively long reconstruction time due to the iterative applications of forward and back projections.

Recently, deep learning approaches have demonstrated impressive performance improvement over conventional iterative methods for low-dose CT and sparse-view CT. The main advantage of deep learning approach over the conventional MBIR approaches is that the network learns the
image statistics in a fully data-driven way rather than using hand-tuned regularizations. While these approaches usually take time for training, real-time reconstruction is possible once the network is trained, making the algorithm very practical in the clinical setting.

While these networks have been designed based on supervised learning technique, in real clinical situation matched low- and routine-dose CT image pairs are difficult to obtain. The matched full/low-dose data are only available when (a) additional full-dose acquisition is available, or (b) simulated low-dose data can be generated from the full-dose acquisition. However, multiple acquisition at different doses is not usually allowed for human study due to the additional radiation dose to patients. Even when such experiments are allowed, the multiple acquisitions are usually associated with motion artifacts due to patients and gantry motions. Therefore, most of the current work uses the simulated low-dose data provided by the vendors [for example, AAPM (American Association of Physicists in Medicine) low-dose grand challenge dataset16]. However, in order to have realistic low-dose images, noise should be added in the sinogram domain, so independent algorithm development without vendor assistance is very difficult. Moreover, there are concerns that the noise patterns in the simulated low-dose image are somewhat different from real low-dose acquisition, so the supervised learning with simulated data can be biased.

To address this unmatched pair problems, Wolterink et al.17 proposed a low-dose CT denoising network with generative adversarial network (GAN) loss so that the distribution of the denoising network outputs can match the routine
dose images. However, one of the important limitations of GAN for CT denoising\cite{17,18,19} is that there is a risk that the network may generate features that are not present in the images due to the potential mode-collapsing behavior of GAN. The GAN mode collapse occurs when the generator network generates limited outputs or even the same output, regardless of the input.\cite{20,21} This happens when GAN is trained to match the data distributions but it does not guarantee that an input and output data are paired in a meaningful way, that is, there can exist various inputs (resp. outputs) that matches to one output (resp. input) despite them generating the same distribution.

In coronary CTA, even though the images at the low-dose and high-dose phases do not match each other exactly due to the cardiac motion, they are from the same cardiac volume so that they have important correspondence. Therefore, one can conjecture that the correctly denoised low-dose phase should follow the routine dose phase image distribution more closely and learning between two phase cardiac images is more effective than learning from totally different images. One of the most important contributions of this work is to show that we can indeed improve the CT images at the low-dose phase by learning the distribution of the images at the high-dose phases using the cyclic consistency by Zhu et al. (cycle GAN)\cite{21} or by Kim et al. (DiscoGAN).\cite{22} Specifically, we train two networks between the two different domains (low dose and routine dose). Then, the training goal is that the two networks should be the inverse of each other.

Thanks to the existence of the inverse path that favors the one to one correspondence between the input and output, the training of the GAN is less affected by the mode collapse.

Furthermore, unlike the classic GAN which generates samples from random noise inputs, our network creates samples from the noisy input that are closely related. This also reduces the likelihood of mode collapse. Another important aspect of the algorithm is the identity loss.\cite{21} The main idea of the identity loss is that a generator \( G: A \mapsto B \) should work as an identity for the target domain image \( y \in B \) such that \( G(y) \approx y \). This constraint works as a fixed-point constraint of the output domain so that as soon as the output signal is generated to match the target distribution, the network no longer changes the signal. Experimental results show that the proposed method is robust to the cardiac motion and contrast changes and does not create artificial features.

2. THEORY

The overall framework of the proposed network architecture is illustrated in Fig. 2. We denote the low-dose CT domain by \( A \) and routine-dose CT domain by \( B \), and the probability distribution for each domain is referred to as \( P_A \) and \( P_B \), respectively. The generator \( G_{AB} \) denotes the mapping from \( A \) to \( B \), and \( G_{BA} \) are similarly defined as the mapping from \( B \) to \( A \). As for the generator, we employ the optimized network for a noise reduction in low-dose CT images in our prior work.\cite{10} In addition, there are two adversarial discriminators \( D_A \) and \( D_B \) which distinguish between measured input images and synthesized images from the generators. Then, we train the generators and discriminators simultaneously. Specifically, we aim to solve the following optimization problem:

\[
\min_{G_{AB}, G_{BA}} \max_{D_A, D_B} \mathcal{L}(G_{AB}, G_{BA}, D_A, D_B).
\]

where the overall loss is defined by:

\[
\mathcal{L}(G_{AB}, G_{BA}, D_A, D_B) = \mathcal{L}_{GAN}(G_{AB}, D_B, A, B) + \lambda \mathcal{L}_{cyclic}(G_{AB}, G_{BA}) + \gamma \mathcal{L}_{identity}(G_{AB}, G_{BA}),
\]

where \( \lambda \) and \( \gamma \) control the importance of the losses, and \( \mathcal{L}_{GAN}, \mathcal{L}_{cyclic} \) and \( \mathcal{L}_{identity} \) denote the adversarial loss, cyclic loss, and identity loss. More detailed description of each loss follows.

2.A. Loss formulation

2.A.1. Adversarial loss

We employ adversarial losses using GAN as proposed in Zhu et al.\cite{21} According to the original GAN,\cite{23} the generator \( G_{AB} \) and discriminator \( D_B \) can be trained by solving the following min–max problem:

\[
\min_{G_{AB}} \max_{D_B} \mathcal{L}_{GAN}(G_{AB}, D_B, A, B) = \mathbb{E}_{x_A \sim P_A}[\log D_B(x_B)] + \mathbb{E}_{x_A \sim P_A}[\log(1 - D_B(G_{AB}(x_A)))],
\]

(3)

where \( G_{AB} \) is trained to reduce a noise in the low-dose CT image \( x_A \) to make it similar to the routine-dose CT image \( x_B \), while \( D_B \) is trained to discriminate between the denoised CT image \( G_{AB}(x_A) \) and the routine-dose CT image \( x_B \). However, we found that the original adversarial loss (3) is unstable during training process; thus, we changed the log-likelihood function to a least square loss as in the least squares GAN (LSGAN).\cite{24} Then, the min–max problem can be changed to the two minimization problems as follows:

\[
\min_{G_{AB}} \mathbb{E}_{x_A \sim P_A}[(D_B(G_{AB}(x_A)) - 1)^2],
\]

(4)

\[
\min_{D_B} \mathbb{E}_{x_B \sim P_B}[(D_B(x_B) - 1)^2] + \frac{1}{2} \mathbb{E}_{x_A \sim P_A}[D_B(G_{AB}(x_A))^2].
\]

(5)

The adversarial loss causes the generator to generate the denoised images that may deceive the discriminator to classify them as the real images at routine doses. At the same time, the adversarial loss will guide the discriminator to well distinguish the denoised image and the routine dose image. Similar adversarial loss is added to the generator \( G_{BA} \), which generates noisy images.

2.A.2. Cyclic loss

With the adversarial losses, we could train the generator \( G_{AB} \) and \( G_{BA} \) to produce the realistic denoised images and
noisy CT images, respectively; but this does not guarantee
that they have an inverse relation described in Fig. 2. To
enable one to one correspondence between the noisy and
denoised image, the cycle which consists of two generators
should be imposed to bring the input \( x_A \) to the original image.
More specifically, the cyclic loss is defined by

\[
\mathcal{L}_{\text{cyclic}}(G_{AB}, G_{BA}) = E_{x_A \sim P_A} \| G_{BA}(G_{AB}(x_A)) - x_A \|_1 + E_{x_B \sim P_B} \| G_{AB}(G_{BA}(x_B)) - x_B \|_1,
\]

(6)

where \( \| \cdot \|_1 \) denotes the \( l_1 \)-norm. Then, the cyclic loss
enforces the constraint that \( G_{AB} \) and \( G_{BA} \) should be inverse of
each other, that is, it encourages \( G_{BA}(G_{AB}(x_A)) \approx x_A \) and
\( G_{AB}(G_{BA}(x_B)) \approx x_B \).

2.A.3. Identity loss

In multiphase CTA, there are often cases where the heart
phase and dose modulation are not perfectly aligned as originally planned. For example, in the multiphase CTA acquisition in Fig. 1, it is assumed that the systolic phase images should be obtained using low-dose modulation, but due to the mismatch with the cardiac cycle from arrhythmia, the systolic phase image noise level may vary and even be in full dose. In this case, the input to the generator \( G_{AB} \) can be at full dose, so it is important to train the generator so that it does not alter such clean images. Similarly, the generator \( G_{BA} \) should not change the input images acquired at the low-dose level. To enforce the two generators, \( G_{AB} \) and \( G_{BA} \), to satisfy these conditions, the following identity loss should be minimized:

\[
\mathcal{L}_{\text{identity}}(G_{AB}, G_{BA}) = E_{x_B \sim P_B} \| G_{BA}(x_B) - x_B \|_1 + E_{x_A \sim P_A} \| G_{AB}(x_A) - x_A \|_1.
\]

(7)

In other words, the generators should work as identity mappings for the input images at the target domain:

\[
G_{AB}(x_B) \approx x_B, \quad G_{BA}(x_A) \approx x_A
\]

(8)

Note that this identity loss is similar to the identity loss for the photo generation from paintings in order to maintain the color composition between the input and output domains. The constraints in Eq. (8) ensure that the correctly generated output images no longer vary when used as inputs to the same network, that is, the target domain should be the fixed points of the generator. As will be shown later in experiments, this constraint is important to avoid creating artificial features.

2.B. Network architecture

The network architecture of two generators, \( G_{AB} \) and \( G_{BA} \), is illustrated in Fig. 3. This architecture is optimized for low-
dose CT image denoising in Kang et al.\textsuperscript{10} To reduce network complexity, images are used directly as inputs to the network instead of the wavelet transform coefficients as in our prior work.\textsuperscript{10} The first convolution layer uses 128 sets of $3 \times 3$ convolution kernels to produce 128 channel feature maps. We have six sets of modules composed of three sets of convolutions, batch normalization, and ReLU layers, and one bypass connection with a ReLU layer. Convolution layers in the modules use 128 sets of $3 \times 3 \times 128$ convolution kernels. In addition, the proposed network has a concatenation layer that concatenates the inputs of each module and the output of the last module, which is followed by the convolution layer with 128 sets of $3 \times 3 \times 896$ convolution kernels. This concatenation layer has a signal boosting effect using multiple signal representations\textsuperscript{10} and provides various paths for gradient backpropagation. The last convolution layer uses 15 sets of $3 \times 3 \times 128$ convolution kernels. Finally, we add an end-to-end bypass connection to estimate the noise-free image while exploiting the advantages of bypass connection in He et al.\textsuperscript{26}

The network architecture of discriminators $D_A$ and $D_B$ is illustrated in Fig. 4. This is from PatchGAN,\textsuperscript{25} which has $70 \times 70$ receptive field and classifies image patches whether they are real or synthesized. Specifically, it consists of five convolution layers including the last fully connected layer. The first convolution layer uses 64 sets of $4 \times 4$ convolution kernels, and the number of convolution kernels in the following layers is twice that of the previous layer except the last fully connected layer. After the last fully connected layer, $5 \times 5$ feature maps are obtained, and we calculate the $l_2$ loss. Arbitrary sized images can be applied to this discriminator network by summing up the $l_2$ loss from each $56 \times 56$ patch, after which the final decision is made.

3. METHODS

3.A. Data: cardiac CT scans

The study cohort comprised 50 CT scans of mitral valve prolapse patients and 50 CT scans of coronary artery disease patients, and the CT scan protocols are described in previous reports.\textsuperscript{27,28} The mean age of the population was 58 ± 13.2 yr, and the mean body weight was 66.2 ± 12.6 kg. Using a second generation dual-source CT scanner (Somatom Definition Flash, Siemens, Erlangen, Germany), electrocardiography (ECG)-gated cardiac CT scanning was performed. Retrospective ECG-gated spiral scan with ECG-based tube current modulation was applied to multiphase of 0–90% of the R–R interval which comprises with a full dose pulsing window of 30–80% of the R–R interval. The tube current was reduced to 20% of the maximum outside the ECG pulsing window\textsuperscript{29} (Fig. 1). A bolus of 70–90 mL of contrast material (Iomeprol, Iomeron 400; Bracco, Milan, Italy) was administered by a power injector (Stellant D; Medrad, Indiana, PA, USA) at a rate of 4.0 mL/s and followed by 40 mL saline. The bolus tracking method (region of interest, the ascending aorta; attenuation threshold level, 100 HU; scan delay, 8 s) was applied to determine scan time.

In all CT scans, tube voltage and the tube current–exposure time product were adjusted according to the patients’ body size, and the scan parameters were as follows: tube voltage, 80–120 kV; tube current–exposure time product, 185–380 mAs; collimation, $128 \times 0.6$ mm; and gantry rotation time, 280 s. Mean effective radiation dose of CCTA was 11.4 ± 6.2 mSv. Standard cardiac filter (B26f) was used for imaging reconstruction.

3.B. Training details

Training was performed by minimizing the loss function (2) with $\lambda = 10$ and $\gamma = 5$. We used the ADAM optimization method to train all networks with $\beta_1 = 0.5$ and $\beta_2 = 0.999$. The number of epochs was 160, which was divided into two phases to control the learning rate during the training. In the first 100 epochs, we set the learning rate to 0.0002, and linearly decreased it to zero over the next epochs.

We performed early stopping at 160 epochs, since the early stopping was shown to work as a regularization.\textsuperscript{30} The size of patch was $56 \times 56$ and the size of mini-batch

| Table I. | Structures selected as diagnostic requirements to assess the diagnostic quality of cardiac images. The structures were evaluated to be sharp with clear visualization. (LCA, left coronary artery; LV, left ventricle; RCA, right coronary artery; RV, right ventricle). |
|----------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Organ    | Structure                                                                                                                              |
|----------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Left/right coronary artery | LCA ostium                                                                                                                             |
|          | LCA distal 1.5 cm                                                                                                                      |
|          | LCA distal                                                                                                                            |
|          | RCA ostium                                                                                                                            |
|          | RCA 1.5 cm                                                                                                                            |
|          | RCA distal                                                                                                                            |
| Cardiac wall | LV septum                                                                                                                             |
| Cardiac cavity | RV free wall margin                                                                                                                  |
|          | LV trabeculation                                                                                                                      |
|          | Left arterial appendage                                                                                                               |
| Aorta    | Aortic root                                                                                                                          |
| Valve    | Aortic valve                                                                                                                          |
|          | Mitral valve                                                                                                                          |

| Table II. Visual grading analysis scores used to evaluate the structure visibility. |
|-----------------|------------------------------------------------------------------------------------------------------------------------------------------|
| Score | Visibility of the structures in relation to the reference images |
| 1   | Poor image quality                                                                                                                   |
| 2   | Lower image quality                                                                                                                  |
| 3   | Mild noise, but acceptable                                                                                                           |
| 4   | Average                                                                                                                             |
| 5   | Good                                                                                                                                |
| 6   | Excellent                                                                                                                            |
FIG. 5. Restoration results from the dataset of mitral valve prolapse patients. Intensity range of the CT image is (−1024, 976) [HU] and the difference image between the input and result is (−150, 150) [HU]. Yellow arrows indicate the distinctly different region between input image from phase 1 and target image from phase 8. [Color figure can be viewed at wileyonlinelibrary.com]

FIG. 6. Restoration results from the dataset of coronary artery disease patients. Intensity range of the CT image is (−924, 576) [HU] and the difference image between the input and result is (−200, 200) [HU]. Yellow arrows indicate the distinctly different region between input image from phase 1 and target image from phase 8. [Color figure can be viewed at wileyonlinelibrary.com]
was 10. Kernels were initialized randomly from a Gaussian distribution.

We have updated the generator and the discriminator at each iteration. We normalized the intensity of the input low-dose CT images and the target routine-dose CT image using the maximum intensity value of the input images, and subtract 0.5 and multiply two to make the input image intensity range as $[-1,1]$. For training, we used 50 cases from the dataset of mitral valve prolapse patients. The proposed method was implemented in Python with the PyTorch and NVIDIA GeForce GTX 1080 Ti GPU was used to train and test the network.

3.C. Evaluation

3.C.1. Visual grading analysis

Image quality was assessed using relative visual grading analysis (VGA). This VGA method is planned to be related to the clinical task to evaluate any structural abnormality that may present at specific anatomical structures in a CT images. Two expert cardiac radiologists established a set of anatomical structures to evaluate image quality. Table I demonstrates the 13 anatomical structures used in this study. The VGA scoring scale are shown in Table II. All CT images including denoising CT images were uploaded on picture archiving and communication system (PACS) for visual grading. Of all, randomly selected 25 CT scans from mitral valve prolapse patients and 25 CT scans from coronary artery disease patients were included for VGA. A total of 1300 CT images (50 selected CT scans $\times$ 13 structures $\times$ original and denoising CT) were scored. Two radiologists performed VGA in consensus, and all CT scans are scored independently, without side-by-side comparison.

3.C.2. Quantitative analysis

The image noise and signal-to-noise (SNR) of all images were obtained at four anatomical structures: ascending aorta, left ventricular cavity, left ventricular septum, and proximal right coronary artery. The size of the region of interest to evaluate SNR was varied to fit each anatomic structure; however, it was confined into each structure without overlapping other structures.

3.D. Statistical analysis

The VGA scores obtained from original CT images and denoising images were compared using Chi-square test. Image noise and SNR were compared using paired $t$-test. $P < 0.05$ indicated statistical significance. Statistical analyses were performed using commercial software (SPSS, Chicago, IL, USA).

4. RESULTS

4.A. Qualitative evaluation

To verify the performance of the proposed method, we tested 50 cases from the dataset of mitral valve prolapse patients and another 50 cases from the dataset coronary...
artery disease patients which were not used in the training session. The results are described in Figs. 5 and 6, respectively. Each row indicates the different patient case, and the restoration results from the first column are shown in the second column. The input low-dose CT images are from phase 1 and the target routine-dose images are from phase 8. Due to the cardiac motion during CT scanning, the shape of the heart and image intensity from the contrast agent are different at the two phases. Distinct differences are indicated by the yellow arrows in the images. The denoised results showed that the proposed method is good at reducing the noise in the input CT images while the texture information and edges are still intact. The difference images showed that the proposed method did not change the detailed information and only removes noise from the input CT images. The proposed method is robust to the type of heart disease as confirmed in another disease cases in Fig. 6. Results showed that the network does not create any artificial features that can disturb the diagnosis while maintaining the crucial information.

We also observed that the proposed method is automatically adapted to the noise levels of the input CT images. Specifically, there are some data that have similar noise levels between phase 1 and phase 8 as shown in Fig. 7. If the input CT images have a noise level similar to the CT target images, we have found that the proposed generator \( G_{AB} \) does not show any noticeable change, as shown in Fig. 7. These results confirm the proposed generator \( G_{AB} \) acts as the identity for the images in the target domain, as shown in Eq. (8).

To compare the performance with the state-of-the-art model-based iterative method (MBIR), we compared our algorithm with the Siemens ADMIRE (Advanced Modeled Iterative Reconstruction) algorithm. ADMIRE is the latest MBIR method from Siemens, which has been improved from SAFIRE (Sinogram Affirmed. Iterative Reconstruction) algorithm. ADMIRE incorporates statistical modeling, both in the raw projection data and in the image domains, such that a different statistical weighting is applied according to the quality of the projection, so ADMIRE is only available for latest scanner (Siemens Flash system). Thus, we cannot provide ADMIRE
images for all patients in our retrospective studies, so we obtained multiphase CTA images from a new patient case.

As shown in Fig. 8, both ADMIRE and the proposed method successfully reduced noise in low-dose CT images. However, the difference images between input and results showed that, in the case of ADMIRE, the edge information was somewhat lost and oversmoothing was observed in the lung region, indicated by red arrows. On the other hand, no structural loss was observed in the proposed method. Moreover, in the left two columns of Fig. 8, we can clearly see the remaining streaking artifacts in the ADMIRE images, while no such artifacts are observed in the proposed method. A similar, consistent improvement by the proposed method was observed in all volume slices.

4.B. Visual grading score and SNR analysis results

All visual scores are significantly higher in denoising CT, representing that the image quality of denoising CT is better \((P < 0.001)\) (Table III). Quantitatively, image noise was decreased, and SNR was significantly increased on denoising CT \((P < 0.05)\) (Table IV, Fig. 9), except no statistically significant SNR changes were detected in the left ventricular cavity where contrast enhanced blood pool was measured by the largest region of interest \((P = 0.96)\).

4.C. Application to AAPM dataset

We have performed additional experiments with AAPM low-dose CT grand challenge dataset which consists of abdominal CT images from 10 patients. We used.

---

**Table III.** Comparison of visual scores between original image and denoising CT image using Chi-square method. (LCA, left coronary artery; LV, left ventricle; RCA, right coronary artery; RV, right ventricle).

| Organ                      | Structure          | Original image | Denoising   | P-value   |
|----------------------------|--------------------|----------------|-------------|-----------|
| Left/right coronary artery | LCA ostium         | 3.5 ± 1.4      | 4.2 ± 1.1   | <0.001    |
|                            | LCA distal 1.5 cm  | 3.4 ± 1.4      | 4.1 ± 1.1   | <0.001    |
|                            | LCA distal 3.0 cm  | 2.3 ± 1.3      | 3.0 ± 1.2   | <0.001    |
|                            | RCA ostium 4.0 cm  | 3.3 ± 1.3      | 4.0 ± 1.0   | <0.001    |
|                            | RCA 1.5 cm         | 3.1 ± 1.6      | 3.9 ± 1.1   | <0.001    |
|                            | RCA distal 2.8 cm  | 2.3 ± 1.4      | 2.8 ± 1.2   | <0.001    |
| Cardiac wall               | LV septum          | 3.1 ± 1.6      | 3.8 ± 1.2   | <0.001    |
|                            | RV free wall margin| 3.2 ± 1.4      | 3.8 ± 1.1   | <0.001    |
| Cardiac cavity             | LV trabeculation   | 3.4 ± 1.5      | 4.3 ± 1.1   | <0.001    |
|                            | Left arterial appendage| 3.0 ± 1.4    | 3.6 ± 0.9   | <0.001    |
| Aorta                      | Aortic root        | 4.3 ± 1.2      | 5.0 ± 0.5   | <0.001    |
| Valve                      | Aortic valve       | 2.8 ± 1.4      | 3.4 ± 0.9   | <0.001    |
|                            | Mitral valve       | 2.8 ± 1.4      | 3.4 ± 1.0   | <0.001    |

**Table IV.** Comparison of standard deviation and signal-to-noise ratio between original CT and denoising CT images measured from selected structures (LV, left ventricle; pRCA, proximal right coronary artery).

|                | Image noise (standard deviation) | SNR                  |
|----------------|----------------------------------|----------------------|
|                | Original image | Denoising | P-value | Original image | Denoising | P-value |
| Ascending aorta| 48.0 ± 26.8 | 39.0 ± 17.0 | 0.003 | 11.1 ± 6.5 | 12.3 ± 6.2 | 0.001 |
| LV cavity      | 69.4 ± 37.9 | 48.5 ± 19.0 | <0.001 | 8.9 ± 12.7 | 9.0 ± 3.5 | 0.96  |
| LV septum      | 65.3 ± 40.8 | 40.9 ± 14.0 | <0.001 | 2.9 ± 1.7 | 3.4 ± 1.3 | 0.015 |
| pRCA           | 70.5 ± 50.7 | 62.1 ± 48.5 | 0.036 | 6.4 ± 3.9 | 7.6 ± 4.4 | 0.034 |

---

![Fig. 9.](image) Standard deviation and signal-to-noise ratio between original CT (red) and denoising CT (purple) images measured from selected structures. (LV, left ventricle; pRCA, proximal right coronary artery). [Color figure can be viewed at wileyonlinelibrary.com]
eight patients’ data for training and validation, and the remaining two patients’ data for the test. In contrast to the existing supervised learning approaches for low-dose CT denosing,\(^9\) here, the training was conducted in an unsupervised manner using the proposed network, with the input and target images randomly selected from the entire dataset. Figure 10(b) showed that the proposed unsupervised learning method provided even better images than the supervised learning, while there are some remaining artifacts in Figs. 10(a) and 10(c). In general, the denoising results by the proposed approach has the competitive denoising performance compared to the supervised learning approach.\(^9\)

### 4.D. Ablation study

To analyze the roles of each building block in the proposed network architecture, we performed ablation studies by excluding the identity loss and/or cyclic loss and using the same training procedures. The results with respect to two different noise levels are illustrated in Figs. 11 and 12, respectively. Recall that the input low-dose CT image in the subfigure (a) and target routine-dose image in the subfigure (e) have different shapes of heart due to the cardiac motion. The results of the proposed method are illustrated in the second column, the results excluding the identity loss are in the third column, and the results excluding the identity loss and cyclic loss are illustrated in the fourth column. We illustrate reconstruction images as well as the difference images between the input and the reconstruction results. We also indicate the artificial features that were not present in the input images by red arrows.

All the reconstruction result images in Fig. 11 show that the noise level is reduced and the edge information is well maintained. In contrast to the proposed method that does not generate any artificial features, the other methods generated some structures which are not present in the

---

Fig. 10. Restoration results from the AAPM challenge dataset using the proposed method and the supervised learning method.\(^9\) Images of (a) the liver; (b) various organs includes the intestine and kidney, etc; (c) the bones. Intensity range of the CT image is \([-300, 300]\) [HU]. [Color figure can be viewed at wileyonlinelibrary.com]
input images. The results excluding the identity loss (third column) are better than the network trained only with GAN loss without including cycle consistency and identity loss (fourth column), but both methods deformed the shape of the heart and removed some structures. Similar observations can be found in Fig. 12 where input CT
image has a similar noise level with target CT image. While the proposed method does not change the original image, the other methods deformed the shape and created the features that were not present in the input image. Considering that artificial features can confuse radiologists in diagnosing the patient’s disease, the result confirmed the critical importance of cyclic loss and the identity loss as proposed by our algorithm.

5. DISCUSSION

Unsupervised learning with GAN has become popular in computer vision literatures, which has demonstrated impressive performance for various tasks, but the classical GAN\textsuperscript{23} using the sigmoid cross entropy loss function is often unstable during training. To address this, we used LSGAN\textsuperscript{24} and the cycle loss.\textsuperscript{21} Convergence plots in Fig. 13 shows that the proposed networks converged stably. Here, $L_{GAN}$ and $L_{DB}$ denoted the loss of generator Eq. (4) and the loss of discriminator Eq. (5). If network reaches the optimal equilibrium state, $L_{GAN}$ and $L_{DB}$ should be reached at 0.25, which was also shown in Fig. 13. The cyclic loss also decreased steadily during training process and converged. This confirms that the network training was well done.

Another critical issue with GAN is the problem of mode collapse. The GAN mode collapse occurs when the generator network generates limited outputs or even the same output, regardless of the input. Unlike the classic GAN, which generates samples from random noise inputs, our network creates samples from the noisy input that are closely related. In addition, the presence of an inverse path reduces the likelihood of mode collapse, and the identity loss prevents the creation of artificial features. Thanks to the synergistic combination of these components of network architectures, the likelihood of mode collapse was significantly reduced, and we have not observed any case where the generated outputs from distinct inputs are the same.

However, there are some limitations of the present studies. The current method mainly focused on multiphase CTA, and the performance of the proposed method is confirmed in this specific application. Also, our training, validation, and test data are generated using the same reconstruction kernel (B26f: cardiac filter). Thus, it is not clear whether our approach can be generalized to different kernels, organs, etc. Even though we provided preliminary results using the AAPM dataset, more extensive study is required to validate the generalizability of the proposed method. These issues are very important for clinical uses, which need to be investigated in separate works.

We agree that once a well-trained network from supervised learning tasks is available, one can use low-dose acquisition for all cardiac phases. However, extensive clinical evaluation is required to have such drastic protocol changes, which is unlikely to happen in the near future. On the other hand, the proposed approach still uses the current acquisition protocols, but provide enhanced images as additional information for radiologists, which can be easily accepted in the current clinical setting. Moreover, in contrast to supervised learning approaches for low-dose CT, the unsupervised learning approaches, such as the proposed one, do not require vendor-supported simulated low-dose data or additional matched full/low-dose acquisition. Therefore, we believe that the potential for the proposed method in terms of science and product development could be significant.

6. CONCLUSION

In this paper, we proposed a cycle-consistent adversarial denoising network for multiphase coronary CT angiography. Unlike the existing supervised deep learning approaches for low-dose CT, our network does not require exactly matched low- and routine- dose images. Instead, our network was designed to learn the image distributions from the high-dose cardiac phases. Furthermore, in contrast to the other state-of-the-art deep neural networks with GAN loss that are prone to generate artificial features, our network was designed to prevent from generating artificial features that are not present in the input image by exploiting the cyclic consistency and identity loss. Experimental results confirmed that the proposed method is good at reducing the noise in the input low-dose CT images while maintaining the texture and edge information. Moreover, when the routine dose images were used as input, the proposed network did not change the images, confirming that the algorithm learns the noise correctly. Radiological evaluation using visual grading analysis scores also confirmed that the proposed denoising method significantly increases the diagnostic quality of the images. Considering the effectiveness and practicability of the proposed method, our method can be widely applied for other CT acquisition protocols with dynamic tube current modulation.
ACKNOWLEDGMENTS

The authors thank Dr. Cynthia MacCollough, the Mayo Clinic, the American Association of Physicists in Medicine (AAPM), and grant EB01705 and EB01785 from the National Institute of Biomedical Imaging and Bioengineering for providing the Low-Dose CT Grand Challenge dataset. This work is supported by Industrial Strategic technology development program (10072064, Development of Novel Artificial Intelligence Technologies To Assist Imaging Diagnosis of Pulmonary, Hepatic, and Cardiac Disease and Their Integration into Commercial Clinical PACS Platforms) funded by the Ministry of Trade Industry and Energy (MI, Korea). This work is also supported by the R&D Convergence Program of NST (National Research Council of Science & Technology) of Republic of Korea (Grant CAP-13-3-KERI).

REFERENCES

1. Hsieh J. Computed Tomography: Principles, Design, Artifacts, and Recent Advances. Bellingham, WA: SPIE; 2009.
2. Beister M, Kolditz D, Kalender WA. Iterative reconstruction methods in X-ray CT. Phys Med. 2012;28:94–108.
3. Ramani S, Fessler JA. A splitting-based iterative algorithm for accelerated statistical X-ray CT reconstruction. *IEEE Trans Med Imaging.* 2012;31:677–688.
4. Sidky EY, Pan X. Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. Phys Med Biol. 2008;53:4777–4807.
5. Leipsic J, LaBounty TM, Heilbron B, et al. Adaptive statistical iterative reconstruction: assessment of image noise and image quality in coronary CT angiography. *Am J Roentgenol.* 2010;195:649–654.
6. Funama Y, Taguchi K, Utsunomiya D, et al. Combination of a low tube voltage technique with the hybrid iterative reconstruction (iDose) algorithm at coronary CT angiography. *J Comput Assist Tomogr.* 2011;35:480–485.
7. Chen G-H, Tang J, Leng S. Prior image constrained compressed sensing (PICCS): a method to accurately reconstruct dynamic CT images from highly undersampled projection data sets. Med Phys. 2008;35:660–663.
8. Renker M, Nance Jr. JW, Schoepf UJ, et al. Evaluation of heavily calcified vessels with coronary CT angiography: comparison of iterative and filtered back projection image reconstruction. *Radiology.* 2011;260:390–399.
9. Kang E, Min J, Ye JC. A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction. *Med Phys.* 2017;44:e360–e375.
10. Kang E, Chang W, Yoo J, Ye JC. Deep convolutional framelet denoising for low dose CT via wavelet residual network. *IEEE Trans Med Imaging.* 2018;37:1358–1369.
11. Chen H, Zhang Y, Kalra MK, et al. Low-dose CT with a residual encoder-decoder convolutional neural network. *IEEE Trans Med Imaging.* 2017;36:2524–2535.
12. Ye JC, Han Y, Cha E. Deep convolutional framelets: a general deep learning framework for inverse problems. *SIAM J Imaging Sci.* 2018;11:991–1048.

"Author to whom correspondence should be addressed. Electronic mail: jong.ye@kaist.ac.kr."