Low-Dose CT Image Reconstruction With a Deep Learning Prior

HYOUNG SUK PARK¹, KYUNGSANG KIM², AND KIWAN JEON¹

¹National Institute for Mathematical Sciences, Daejeon 34047, South Korea
²Gordon Center for Medical Imaging, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114, USA

Corresponding author: Kiwan Jeon (jeonkiwan@nims.re.kr)

This work was supported by the National Institute for Mathematical Sciences (NIMS) grant funded by the Korean government under Grant NIMS-B20900000.

ABSTRACT In low-dose computed tomography (LDCT), a penalized weighted least squares (PWLS) approach that incorporates the Poisson statistics of X-ray photons can significantly reduce excessive quantum noise. To improve the quality of LDCT images, prior information such as the total variation, Markov random field, and nonlocal mean, can be imposed onto the target image. However, this information may be limited to reflect the characteristics of the target images, thereby resulting in unexpected side effects (e.g. blurry images). In this paper, we propose a PWLS method combined with a deep learning prior, which is learned from standard dose CT (SDCT) images. The proposed model learns a noise reduction function that maps an LDCT image to its corresponding SDCT images and estimates the prior distribution using a Pearson $\chi^2$ divergence. The model can be converted to the least squares generative adversarial network with an added PWLS objective, where the optimal generator acts as the noise reduction function. We reformulate the proposed model as a constrained optimization problem and solve it using the alternating optimization (AO) algorithm. Clinical SDCT and simulated LDCT scans of ten patients were used to show the validity of the proposed method. Results show that the proposed method outperforms other PWLS methods, by imposing priors such as the total variation and the nonlocal mean.

INDEX TERMS Low-dose computed tomography, penalized weighted least squares, deep learning, denoising.

I. INTRODUCTION

X-ray computed tomography (X-ray CT) is a diagnostic imaging modality widely used in medical radiology. It produces cross-sectional images of the human body with high resolution. However, with the growing concerns regarding the risk of radiation induced cancer, there is an emergent need for radiation dose reduction [7], [16]. In clinical practice, decreasing X-ray current is the most common method of reducing the X-ray dose [15]. The quality of low-dose CT (LDCT) is partly limited by excessive quantum noise, which results in the degradation of diagnostic performance.

Over the last few decades, numerous iterative reconstruction (IR) methods for noise reduction have been proposed. Some of these methods have incorporated statistical methods such as Maximum a Posteriori (MAP) approach for data fitting in a sinogram space. By assuming that X-ray photons follow a Poisson distribution on reaching an X-ray detector, the MAP model can be reduced to a penalized weighted least squares (PWLS) optimization, including a term relating to the prior distribution of the target images [17], [20]. Owing to the difficulty in calculating the prior distribution, alternative priors reflecting the spatial correlation among neighboring pixels, such as total variation (TV) [6], [23], fractional-order TV [25], nonlocal TV [18], Markov random fields theory [17], [20], [21], or nonlocal means (NLM) [3], [10], [12], [26], have been imposed onto the target image. Although the IR method improves the overall image quality, some issues remain to be solved. First, designing a prior that can effectively reflect the characteristics of the target image continues to be a challenge. For example, TV, NLM, and their variants are sensitive to noise, which can lead to over smoothing or cartoon-like textures. In addition, the manual selection of internal parameters for TV and NLM is difficult but a crucial factor influencing the denoising performance. Second, the IR method is computationally expensive.
Recently, Park et al. [19] have proposed an image-based deep learning model that can learn a noise reduction function that maps LDCT images to their standard dose CT (SDCT) image counterparts by imposing prior onto the the target images. Using several SDCT samples, they estimated the SDCT image distribution (i.e., prior distribution) in the sense of Kullback-Leibler (KL) divergence. Their method allowed the problem to be expressed in a generative adversarial network (GAN) framework. The results showed that their approach can infer the desired SDCT-like image priors based on the SDCT images used for training. In addition, no loss function between the LDCT images and their corresponding SDCT images was required. Hence, training using unpaired LDCT and SDCT datasets was proven to be feasible. However, the approach assumed the presence of Gaussian noise in the CT images, which may not be valid in clinical practice.

In this paper, we propose a deep PWLS method by combining it with a deep learning prior. Unlike the previous work of [19], we estimate the prior distribution in the sense of Pearson χ² divergence, using several SDCT samples. We then convert the proposed model into a least squares-GAN (LS-GAN) [13] that adopts the least squares loss function for the discriminator, including a PWLS objective. Compared to the Jensen-Shannon or KL divergences, Pearson χ² divergence is known to provide a more stable learning process as it has no zero forcing properties [13], [14]. Our proposed model is then reformulated as a constrained optimization problem and solved using the alternating optimization (AO) algorithm. While training for the noise reduction function is performed in an iterative way, the corrected image is obtained directly from the trained noise reduction function. Numerical simulations are performed to confirm the validity of the proposed method.

II. METHOD
Let y ∈ ℝ^M be a measured X-ray photon counts reaching to detector, and let x ∈ ℝ^N be an unknown attenuation coefficient to be determined. CT system tries to recover x from measured y. Define A = [aij] ∈ ℝ^M×N by forward operator whose element aij denotes the effective intersection length of projection line j with voxel i. Considering the monochromatic X-ray, the expected number of measured X-ray photons at detector j, denoted by ỹ_j, can be simply modeled as

\[ ỹ_j = l^0 \exp \left( -\sum_{i=1}^{N} a_{ij} x_j \right), \quad 1 \leq i \leq M, \]  

where l^0 is the number of X-ray photons generated at source.

We start from maximum a-posteriori (MAP) approach. Given y, find \( \hat{x} \) that maximizes the conditional probability \( p_{xy}(x|y) \):

\[ \hat{x} = \arg \max_{x} p_{xy}(x|y) = \arg \max_{x} \log p_{x}(x) + \mathcal{L}(y|x), \]  

where \( \mathcal{L}(y|x) = \log p_{y|x}(y|x) \). Here, the distribution \( p_{x}(x) \) is generally unknown. In the following section, we estimate \( p_{x}(x) \) using many noise-free images \( \{x_i\}_{i=1}^{K} \) sampled from the \( p_{x}(x) \).

A. ESTIMATION OF PRIOR USING DEEP LEARNING
Let z and x be a LDCT and SDCT image reconstructed from low-dose scan \( y_l \) and standard-dose scan \( y_s \), respectively. Regarding x as a noise-free image, let \( G : z \rightarrow x \) be a noise reduction function. For notational convenience, we write \( y = y_l \). Together with many samples \( \{z_i\}_{i=1}^{K} \) and \( \{x_i\}_{i=1}^{K} \), we hope to find \( G \) maximizing the expectation of \( \log p_{x(G(z))}(G(z)) \), which can be expressed as:

\[ G_{\text{MAP}} := \arg \min_{\theta_{G}} E_{z \sim p_{z}} \left[ -\log(p_{x}(G(z; \theta_{G})) - \mathcal{L}(y|G(z)) \right] \]

\[ = \arg \min_{\theta_{G}} E_{G(z) \sim p_{G|z}} \left[ \log(p_{x}(G(z; \theta_{G}))) \right] \]

\[ -E_{z \sim p_{z}} \left[ \mathcal{L}(y|G(z; \theta_{G})) \right] = \arg \min_{\theta_{G}} H(p_{G|z}(p_{x}) - E_{z \sim p_{z}} \left[ \mathcal{L}(y|G(z; \theta_{G})) \right], \]  

where \( H(x, y) \) is the cross-entropy of \( x \) and \( y \). However, due to the difficulty in handling \( H(p_{G(z)}, p_{x}) \) only from the samples \( \{G(z_i)\}_{i=1}^{K} \), we instead estimate \( p_{x} \) by using the Pearson \( \chi^{2} \) divergence \( \chi^{2}_{\text{Pearson}}(p_{G|z} + p_{z}{2p_{x}}{2}) \). This quantity is zero when \( p_{G(z)} = p_{x} \). Now, the problem (3) can be reduced to

\[ G_{s} := \arg \min_{\theta_{G}} \chi^{2}_{\text{Pearson}}(p_{G|z} + p_{z}{2p_{x}}{2}) - E_{z \sim p_{z}} \left[ \mathcal{L}(y|G(z; \theta_{G})) \right]. \]

It can be rewritten with adversarial loss [13], [14], which is given by

\[ G_{s} = \arg \min_{\theta_{G}} E_{z \sim p_{z}} \left[ D(G(z; \theta_{G})) - \mathcal{L}(y|G(z; \theta_{G})) \right], \]  

where \( D \) is the optimum of the following problem.

\[ D = \arg \min_{\theta_{D}} E_{x \sim p_{x}} \left[ (1 - D(x; \theta_{D}))^2 \right] + E_{z \sim p_{z}} \left[ (1 + D(G(z; \theta_{D}))^2 \right]. \]

The proof of (5) is the same as that of Theorem 1 in the paper [14]. The only difference lies in the use of the term \( E_{z \sim p_{z}} \left[ \mathcal{L}(y|G(z; \theta_{G})) \right]. \) For more details, please refer to Appendix IV. The problem in (5) can be viewed as the Least-squares GAN with an additional constraint term, and hence the function \( G \) can be learned with the training samples \( \{z_{k}\}_{k=1}^{K} \) and \( \{x_{k}\}_{k=1}^{K}. \) The least squares loss for the discriminator in (6) allows the proposed method to have a stable performance during training [14], [27].

B. OPTIMIZATION METHOD
We optimize the problem (5) and (6) in alternating steps. For fixed \( D \), we first optimize the problem (5). To decouple the denoising function \( G \) from the term \( \mathcal{L} \), we convert the problem (5) to the following constraint format:

\[ \arg \min_{\hat{x}} E_{z \sim p_{z}} \left[ D(\hat{x})^2 - \mathcal{L}(y|\hat{x}) \right] \]

s.t. \( \hat{x} = G(z; \theta_{G}). \)
We now plug this problem into the quadratic penalty formulation [5], [22].

$$\arg\min_{\tilde{x}, \theta_g} E_{z \sim p_z} \left[ D(\tilde{x})^2 + L(y|\tilde{x}) + \frac{\rho}{2} \|\tilde{x} - G(z; \theta_g)\|^2 \right]$$  \hspace{1cm} (8)

with a sufficiently large penalty parameter $\rho$. This problem can be solved iteratively using AO algorithm in three steps: for each sample $z \sim p_z(z)$,

$$\tilde{x}^{(n+1)} = \arg\min_{\tilde{x}} D(\tilde{x})^2 + L(y|\tilde{x}) + \frac{\rho^{(k)}}{2} \|\tilde{x} - G(z; \theta_g^{(n)})\|^2,$$

$$\theta_g^{(n+1)} = \arg\min_{\theta_g} E_{z \sim p_z} \left[ \|\tilde{x}^{n+1} - G(z; \theta_g)\|^2 \right],$$

$$\rho^{(k+1)} = C \times \rho^{(k)},$$ \hspace{1cm} (11)

where $C$ is some constant. Assuming that $y_i$'s are mutually independent and follow the Poisson distribution, $L(y|\tilde{x})$ in (9) can be expressed as

$$L(y|\tilde{x}) = \sum_{i=1}^{M} \log p(y_i|\tilde{x}) = \sum_{i=1}^{M} \log p(y_i|\tilde{y}_i) = \sum_{i=1}^{M} y_i \log \tilde{y}_i - \tilde{y}_i - \log y_i!.$$ \hspace{1cm} (12)

Second order Taylor’s expansion of $L(y|\tilde{x})$ provides that [20]

$$L(y|\tilde{x}) = -\frac{1}{2} \|A\tilde{x} - p\|^2_W,$$ \hspace{1cm} (13)

where $p \in \mathbb{R}^M$ is the projection data whose element $p_i = -\log(y_i/\beta_i)$, $W$ is a diagonal matrix whose elements $w_i = y_i$. Here, $\|x\|^2_W$ is the weighted norm defined as $\|x\|^2_W = x^T W x$. Substituting (13) into (9), we have

$$\tilde{x}^{(n+1)} = \arg\min_{\tilde{x}} D(\tilde{x})^2 + \frac{1}{2} \|A\tilde{x} - p\|^2_W + \frac{\rho^{(k)}}{2} \|\tilde{x} - G(z; \theta_g^{(n)})\|^2.$$ \hspace{1cm} (14)

Due to the difficulty in handling the network term $D(\tilde{x})^2$, we use the gradient-descent method to solve the problem (14).

$$\tilde{x}^{(n+1)} = \tilde{x}^{(n)} - \eta_k \left[ \nabla_{\tilde{x}} D(\tilde{x})^2 + A^T W (A\tilde{x}^{(n)} - p) + \rho (\tilde{x}^{(n)} - G(z; \theta_g^{(n)}) \right],$$ \hspace{1cm} (15)

where $\eta_k$ is the step size. In our study, we empirically choose it so that algorithm converges.

The problem (10) is the training problem of $G$ with paired dataset $\{z_i, x_i^{n+1}\}_{i=1}^K$, and its update is performed by using first-order optimization method (i.e., Adam optimization [11]).

Now, for fixed $G$, we solve the problem (6). The update of $\theta_d$ is performed by using the Adam optimization, and is performed alternatively with the update of $\theta_g$ to improve the stability of learning algorithm [8]. The overall algorithm of the proposed method is described in Algorithm 1.

C. IMPLEMENTATION DETAILS

Solving the problem of (14) is challenging due to the non-linear network term. We use the gradient-descent method as in (15), where the derivative term $\nabla_x D(x)^2$ is implemented as follows. Let $F$ be a loss function given by $F(b) = D(x + b)^2$, where $b$ is trainable variable. Then, the gradient descent method provides

$$b^{(1)} = b^{(0)} - \eta_b \nabla_b D(x + b)^2 = b^{(0)} - \eta_b \nabla_b D(x + b)^2.$$ \hspace{1cm} (16)

The last equality follows from the chain rule. For $b^{(0)} = 0$ and $\eta_b = -1$, we have

$$b^{(1)} = \nabla_x D(x)^2.$$ \hspace{1cm} (17)

This means that the derivative term $\nabla_x D(x)^2$ can be easily implemented by doing the standard backpropagation of $F(b)$ once with the choice of $b^{(0)} = 0$ and $\eta_b = -1$. Fig. 1 shows the result of the gradient-descent method for the problem of minimizing $D(x)^2$, where the derivative of $D(x)^2$ is implemented as in (17). As shown in Fig. 1, the $D(x)^2$ converges to local minimum as the number of iterations increases.

In our implementation, the initial weights $\theta_g^{(0)}$ and $\theta_d^{(0)}$ are respectively chosen as the solutions of the image-based deep learning (IM-DL) model given by

$$\theta_{g}^{(0)} = \arg\min_{\theta_g} E_{z \sim p_z} \left[ D(G(z; \theta_g))^2 + \alpha \|G(z; \theta_g) - z\|^2 \right],$$ \hspace{1cm} (18)

and

$$\theta_{d}^{(0)} = \arg\min_{\theta_d} E_{x \sim p_x} \left[ (1 - D(x; \theta_d))^2 \right] + E_{z \sim p_z} \left[ (1 + D(G(z); \theta_d))^2 \right].$$ \hspace{1cm} (19)

where parameter $\alpha = 10$ as in [19], and training was run 200 epochs using Adam optimization. Unlike the proposed method, the IM-DL model contains the least squares term $\|G(z; \theta_g) - z\|^2$. We solve the PWLS problem using the

**FIGURE 1.** The result of the gradient-descent method for $D(x)^2$ over the number of iterations.
gradient-descent method for 2000 iterations with a step size of 1e-6, and used it as initial $x^{(0)}$. In our implementation, the number of outer iterations ($OutIter$), inner iterations ($InIter$), and epochs ($Epoch$) were set to 20, 100, and 10, respectively. The parameter $C$ in (11) was set to $C = 2^{8\log_{2}9}$, so that $C_{OutIter} = 2^{8}$. 

D. NETWORKS ARCHITECTURES & TRAINING

For generator (or noise reduction function), we adopt the deep convolutional framelets [24]. It is a multi-scale convolutional network (CNN) that is composed of encoder and decoder paths with skipped and concatenation connections between them. The encoder and decoder paths are respectively composed of three stages. Each stage consists of two repeated convolutions (conv) with a $3 \times 3$ window. Each convolution is followed by a batch normalization (bnorm) and a leaky rectified linear unit (LeReLU). 2-D Haar wavelet decomposition (wave-dec) and recomposition (wave-rec) are applied for downsampling and upsampling of the features, respectively. High pass filters after wavelet decomposition skip directly to the decoder path at same stage, while low pass filters (LF) are concatenated with the features in the decoder path. At the end of decoder path, an $1 \times 1$ convolution layer is added to generate a output with single channel.

For discriminator we adopted the standard CNN that consists of three convolution layers with a $4 \times 4$ window and strides of two in each direction of the domain. Each layer is followed by a batch normalization and a leaky ReLU with a slope of 0.2. At the end of the architecture, a $1 \times 1$ convolution layer is added to generate 1-dimensional output data. The overall architecture of the proposed method is shown in Fig. 2. In Fig. 2, the parameter $n_{g}$ and $n_{d}$ denote the the initial number of filters of generator and discriminator, respectively. Here, we set to $n_{g} = 64$ and $n_{d} = 64$.

In our simulations, the generator/discriminator training was performed using the Adam solver [11] with a learning rate of 0.0002, and mini-batch size of 1. Training was implemented using Tensorflow [1] on a CPU (Intel(R) Core(TM) i9-9980XE, 3.0GHz) and a GPU (NVIDIA, Titan RTX 24GB) systems. The network weights were initialized following a Gaussian distribution with a mean of 0 and standard deviation of 0.01. No data normalization and augmentation were performed during training.

III. RESULTS

We used the datasets provided by “the 2016 NIH-AAPM-Mayo Clinic Low Dose CT Grand Challenge”. Clinical CT scans of ten patients were included, and these scans were acquired using commercial CT scanners (Somatom Definition AS+, or Somatom Definition Flash operated in single-source mode, Siemens Healthcare, Germany) with a reference tube voltage of 120 kVp and a current of 200 mA. Quarter-dose CT (i.e., LDCT) scans were generated by adding Poisson quantum noise to the clinical SDCT scans. All the projection data or sinograms obtained from the helical CT scanners using curve-detectors were rebinned to the fan beam projection data with flat-panel detectors. Standard filtered-backprojection [2] with Ram-Lak filter was applied to reconstruct the LDCT and SDCT images of size $512 \times 512$. Of the total ten scans, we used eight for training and the remaining two for testing. From the training dataset, we randomly extracted 30 pairs of LDCT and SDCT images, and they were used as validation dataset for the hyperparameter tuning of the proposed model.

The proposed method was compared with the PWLS method with total variation regularization (PWLS-TV) [23], nonlocal mean (PWLS-NLM) [10], IM-DL [19], and IM-DL with the cycle consistency loss (IM-Cycle) [9]. The model of PWLS-TV and PWLS-NLM is given by

$$\arg \min_{x} \mathcal{R}(x) + \alpha\|Ax - y\|_{W}^{2},$$

where $\mathcal{R}(x)$ is the regularization terms for TV [23] and NLM [10], respectively. The hyper parameters of $\alpha = 50$ and 25 for
PWLS-TV and PWLS-NLM, which were manually selected with the minimum MSE in our experiments. IM-DL model for image denoising is given by

$$\arg \min_{\theta_g} \chi^2_{\text{Pearson}}(p_G(z) + px^2\|px)$$

$$+ \alpha E_{z \sim p_z} \left[ \|G(z; \theta_g) - z\|^2 \right].$$  \hspace{1cm} (21)

Here, for fair comparison, we used the $\chi^2_{\text{Pearson}}$ divergence, instead of KL divergence.

The IM-Cycle model adopts the Cycle-GAN [27] for image denoising. In our framework, the optimal generator G can be viewed as the solution of the following problem:

$$\arg \min_{\theta_g, \theta_r} \chi^2_{\text{Pearson}}(p_G(z) + px^2\|px)$$

$$+ \alpha E_{z \sim p_z} \left[ \|G(z; \theta_g) - z\|^2 \right]$$

$$+ \gamma E_{x \sim p_x} \left[ \|T(x; \theta_t) - x\|^2 \right]$$

$$+ \gamma E_{x \sim p_x} \left[ \|G(T(x; \theta_t); \theta_g) - x\|^2 \right].$$  \hspace{1cm} (22)

where $T : x \rightarrow z$ is the inverse function that maps $z$ into $x$. For fair comparison, we constructed the same architectures (i.e., both $G$ and $D$) as the proposed method. Note that the difference between IM-Cycle and IM-DL is that it use the cycle-consistency loss (the last two terms in (22)). As suggested in [9], [19], the parameters in IM-DL and IM-Cycle were set to $\alpha = 10$ $(\alpha, \gamma) = (10, 5)$, respectively.

Fig.3 shows the graph of the objective function of (7) over the number of outer iterations for AAPM dataset. As can be seen, the objective values converge stably to a local minimum during training.

Figs. 4 and 5 compare the results of the proposed method with the those of conventional methods such as PWLS-TV, PWLS-NLM, IM-Cycle and IM-DL methods using the AAPM test datasets (patient of L506 and L096). To clearly demonstrate the denoising effect, the magnified regions-of-interests (ROIs) are marked as rectangular regions in the LDCT images. As shown in Figs. 4 and 5, the LDCT images suffer from the noise streaking artifacts due to the low radiation dose. The metastatic legion (marked by arrow in Fig. 5) is barely seen in the LDCT image. All the denoising methods significantly reduced the noise streaking artifacts in the LDCT images, while improving lesion visibility. However, PWLS-TV and PWLS-NLM methods resulted in quite blurry images owing to the inherent nature of TV and NLM, while the proposed method clearly recovered the morphological structures of tissues in the body, as shown in Fig. 4. The denoised images reconstructed by the IM-DL, IM-Cycle and the proposed method are visually more similar to the SDCT images than those reconstructed by the PWLS-TV and PWLS-NLM methods. This is because the presented deep
learning methods exploit the characteristics of SDCT images as a prior. Fig. 6 shows comparison among the deep learning methods for the LDCT image of the pelvis. Insufficient number of photons reaching to detectors results in several noise streaking artifacts along the direction of greatest attenuation. The proposed method more clearly reduces noise streaking artifacts than IM-DL and IM-Cycle. For an objective evaluation of noise level, we calculated the standard deviation (SD) values on a marked ROI (see Fig. 6), and the proposed method achieved the best SD value. The IM-DL and IM-Cycle only used the pair of LDCT and SDCT images during the training procedure, whereas the proposed method used the images corrected from the sinogram, as well as the pair of LDCT and SDCT images. Hence, the proposed method showed the potential to reduce severe noise streaking artifacts even more than the SDCT images.

For performance comparison, we computed the average peak signal to noise ratio (PSNR), structure similarity (SSIM), and mean square error (MSE) between the corrected images and their SDCT images for the test AAPM datasets. The averaged quantities are summarized in Table 1. The results showed that the proposed method achieves the best PSNR, SSIM, and MSE values, whereas the PWLS-NLM method shows the worst performance in terms of PSNR.
and MSE. IM-Cycle method shows better performance than IM-DL method. The added cycle consistency loss leads to performance improvement in denoising task, whereas it increases training time by more than two times. As suggested in [9], [19], training of the both IM-DL, IM-Cycle methods can be performed in a patch-by-patch manner, which can also lead to an improvement in the performance of the IM-DL and IM-Cycle methods.

IV. DISCUSSION AND CONCLUSION

In this paper, we proposed a PWLS method by combining it with a deep learning prior to improve the image quality in LDCT. We estimated the prior distribution of the target images in the sense of Pearson $\chi^2$ divergence using many SDCT samples. The proposed model can be converted into a LS-GAN with a PWLS objective, where the optimal generator is the denoising function that maps the LDCT images to their corresponding SDCT images. We reformulate the model as constraint problem and use the AO method to decouple the denoising function and the penalized weighted objective. By doing so, the standard backpropagation algorithm provided by Tensorflow can be used to update the weight of the denoising function.

Numerical experiments show that the proposed method clearly reduces the quantum noise in LDCT images, while preserving the morphological structures of the human body. The proposed method outperforms the PWLS-TV, PWLS-NLM, IM-Cycle and IM-DL methods, in terms of quantitative metrics such as MSE, SSIM, and PSNR (Table 1). TV and NLM are parametric regularizers, and the reconstructed image quality depends on the experimental setup. For a fair comparison, TV and NLM hyperparameters with the minimum MSEs were manually selected. The PWLS-TV and PWLS-NLM methods produce blurry images but seem to provide better lesion detectability than the proposed method (or even than the SDCT images). The proposed method recovers the images with a deep image prior learned from the SDCT images used for training. Depending on the quality of the SDCT images, the ability to detect lesion can be improved. Unlike the IM-DL and IM-Cycle, the proposed approach incorporates Poisson statistics; hence, the proposed method can reduce the noise streaking artifacts more effectively due to the quantum noise, as shown in Fig. 6. For training, IM-DL and IM-Cycle methods only require CT images, whereas the proposed method requires both sinograms and CT images. The training procedure of the proposed method is more complicated because it is performed with multiple nested loops.

For the AAPM training datasets of size $512 \times 512$, it took approximately 7 days to train the proposed network. Nevertheless, the correction results can be obtained directly through the trained denoising function (or generator), and hence the execution time of the proposed method is much shorter than those of the model-based approaches such as PWLS-TV and PWLS-NLM. The proposed method and the PWLS-TV method require approximately 50 milliseconds and 16 seconds to obtain a single corrected image of size $512 \times 512$ on a GPU system, respectively.

The performance of the proposed method needs further improvement. There is no convergence theory of the AO method used in the proposed method owing to the non-convexity of the two neural networks: generator and discriminator. We empirically showed that the objective function in (7) converges with the increase in the number of iterations (Fig. 3). In addition, we empirically chose the hyperparameters and iteration numbers of the proposed method based on the experiments. For example, Table 2 illustrates the performance evaluation of the proposed method on the validation dataset for different numbers of filters and different sets of
iteration parameters. In this study, we manually chose parameters with the minimum MSE. However, the chosen values may be not optimal. Further investigation will be needed to find more efficient parameters for practical use. Some studies [4], [19] have investigated the robustness of learning-based denoising methods for noise levels different from the one used in training, showing that the denoising performance of the learning-based methods can be degraded for datasets with different noise levels. The performance can be improved if training is performed together with the LDCT images acquired at different noise levels [4]. The validity of the proposed method should be demonstrated for actual clinical use, which is a part of our future work. Furthermore, we will focus on broadening the scope of the proposed method to provide a solution to other tomography reconstruction problems, such as positron emission tomography and magnetic resonance imaging.

APPENDIX

In this Appendix, we prove the equation (5). Let $J_d(D)$ and $J_g(G)$ be functions given by

$$J_d(D) = E_{x \sim p_x} \left[ (1 - D(x))^2 \right] + E_{z \sim p_z} \left[ (1 + D(G(z)))^2 \right] \tag{23}$$

and

$$J_g(G) = E_{x \sim p_x} \left[ D(x)^2 \right] + E_{z \sim p_z} \left[ D(G(z))^2 \right] - E_{z \sim p_z} \left[ \mathcal{L}(y|G(z)) \right] \tag{24}$$

Given a fixed $G$, we first minimize $J_d(D)$. Writing both $G(z)$ and $x$ to variable $\omega$, optimal $D_{opt}$ can be computed as

$$D_{opt}(\omega) = \frac{p_x(\omega) - p_G|z(\omega))}{p_x(\omega) + p_G(\omega)} \tag{25}$$

By plugging $D_{opt}$ into $J_g(G)$, we have

$$J_g(G) = E_{x \sim p_x} \left[ D_{opt}(x)^2 \right] + E_{z \sim p_z} \left[ D_{opt}(G(z))^2 \right] - E_{z \sim p_z} \left[ \mathcal{L}(y|G(z)) \right] \tag{26}$$

$$= \int \left[ \frac{(p_x(\omega) - p_G|z(\omega))^2}{p_x(\omega) + p_G(\omega)} \right] d\omega - E_{z \sim p_z} \left[ \mathcal{L}(y|G(z)) \right] \tag{27}$$

$$= \chi^2_{Pearson}[p_G(z) + p_x] - \chi^2_{Pearson}[p_G(z)], \tag{28}$$

where $\chi^2_{Pearson}(p, q) := \int \frac{(p(x) - q(x))^2}{q(x)} dx$. Since the first term in (24) is independent of $G$, minimizing $J_g(G)$ is equivalent to minimizing the equation (5), and this completes the proof.

REFERENCES

[1] M. Abadi, “TensorFlow: Large-scale machine learning on heterogeneous distributed systems,” 2016, arXiv:1603.04467. [Online]. Available: http://arxiv.org/abs/1603.04467
[2] R. N. Bracewell and A. C. Riddle, “Inversion of fan-beam scans in radio astronomy,” Astrophys. J., vol. 150, p. 427, Nov. 1967.
[3] Y. Chen, D. Gao, C. Nie, L. Luo, W. Chen, X. Yin, and Y. Lin, “Bayesian statistical reconstruction for low-dose X-ray computed tomography using an adaptive-weighting nonlocal prior,” Comput. Med. Imag. Graph., vol. 33, no. 7, pp. 495–500, Oct. 2009.
[4] H. Chen, “Low-dose CT via convolutional neural network,” Biomed. Opt. Express, vol. 8, no. 2, pp. 679–694, 2017.
[5] R. Courant, “Variational methods for the solution of problems with equilibrium and vibration,” B. Amer. Math. Soc., vol. 49, pp. 1–23, 1943.
[6] X. Dong, T. Niu, and L. Zhu, “Combined iterative reconstruction and image-domain decomposition for dual energy CT using total-variation regularization,” Med. Phys., vol. 41, no. 5, Apr. 2014, Art. no. 051909.
[7] A. B. de González and S. Darby, “Risk of cancer from diagnostic X-rays: Estimates for the UK and 14 other countries,” Lancet, vol. 363, no. 9406, pp. 345–351, Jan. 2004.
[8] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” in Proc. NIPS, 2014, pp. 2672–2680.
[9] E. Kang, H. J. Koo, D. H. Yang, J. B. Seo, and J. C. Ye, “Cycle consistent adversarial denoising network for multiphase coronary CT angiography,” Med. Phys., vol. 46, no. 2, pp. 550–562, Feb. 2019.
[10] K. Kim, G. El Fakhri, and Q. Li, “Low-dose CT reconstruction using spatially encoded nonlocal penalty,” Med. Phys., vol. 44, no. 10, pp. 376–390, Oct. 2017.
[11] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” 2014, arXiv:1412.6980. [Online]. Available: http://arxiv.org/abs/1412.6980
[12] J. Ma, H. Zhang, Y. Gao, J. Huang, Z. Liang, Q. Feng, and W. Chen, “Iterative image reconstruction for cerebral perfusion CT using a pre-contrast scan induced edge-preserving prior,” Phys. Med. Biol., vol. 57, no. 22, pp. 7519–7542, Nov. 2012.
[13] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. Paul Smolley, “Least squares generative adversarial networks,” 2016, arXiv:1611.04076. [Online]. Available: http://arxiv.org/abs/1611.04076
[14] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, “On the effectiveness of least squares generative adversarial networks,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 41, no. 12, pp. 2947–2960, Dec. 2019.
[15] L. F. McCaig and C. W. Burt, “National Hospital Ambulatory Medical Care Survey: 2002 emergency department summary,” Adv. Data, vol. 340, pp. 1–34, Jun. 2004.
[16] D. L. Miglioretti, E. Johnson, A. Williams, R. T. Greenlee, S. Weinmann, L. I. Solberg, H. S. Feigelson, D. Roblin, M. J. Flynn, N. Vanneman, and R. Smith-Bindman, “The use of computed tomography in pediatrics and the associated radiation exposure and estimated cancer risk,” JAMA Pediatr., vol. 167, no. 8, pp. 700–707, 2013.
[17] T. Li, X. Li, J. Wang, J. Wen, H. Lu, J. Hsieh, and Z. Liang, “Nonlinear sinogram smoothing for low-dose X-ray CT,” IEEE Trans. Med. Imag., vol. 35, no. 12, pp. 2505–2513, Oct. 2004.
[18] J. Liu, H. Ding, S. Molloi, X. Zhang, and H. Gao, “TCMR: Total image constrained material reconstruction via nonlocal total variation regularization for spectral CT,” IEEE Trans. Med. Imag., vol. 35, no. 12, pp. 2578–2586, Dec. 2016.
[19] H. S. Park, J. Baek, S. K. You, J. K. Choi, and J. K. Seo, “Unpaired image denoising using a generative adversarial network in X-ray CT,” IEEE Access, vol. 7, pp. 110414–110425, 2019.
[20] K. Sauer and C. Bouman, “A local update strategy for iterative reconstruction from projections,” IEEE Trans. Signal Process., vol. 41, no. 2, pp. 534–548, Dec. 1993.
[21] J. Wang, T. Li, H. Lu, and Z. Liang, “Penalized weighted least-squares approach to sinogram noise reduction and image reconstruction for low-dose X-ray computed tomography,” *IEEE Trans. Med. Imag.*, vol. 25, no. 10, pp. 1272–1283, Oct. 2006.

[22] Y. Wang, J. Yang, W. Yin, and Y. Zhang, “A new alternating minimization algorithm for total variation image reconstruction,” *SIAM J. Imag. Sci.*, vol. 1, no. 3, pp. 248–272, Jan. 2008.

[23] Q. Xu, X. Mou, G. Wang, J. Sieren, E. A. Hoffman, and H. Yu, “Statistical interior tomography,” *IEEE Trans. Med. Imag.*, vol. 30, no. 5, pp. 1116–1128, May 2011.

[24] J. C. Ye, Y. Han, and E. Cha, “Deep convolutional framelets: A general deep learning framework for inverse problems,” *SIAM J. Imag. Sci.*, vol. 11, no. 2, pp. 991–1048, Jan. 2018.

[25] Y. Zhang, Y. Wang, W. Zhang, F. Lin, Y. Pu, and J. Zhou, “Statistical iterative reconstruction using adaptive fractional order regularization,” *Biomed. Opt. Express*, vol. 7, no. 3, pp. 1015–1029, 2016.

[26] H. Zhang, D. Zeng, J. Lin, H. Zhang, Z. Bian, J. Huang, Y. Gao, S. Zhang, H. Zhang, Q. Feng, Z. Liang, W. Chen, and J. Ma, “Iterative reconstruction for dual energy CT with an average image-induced nonlocal means regularization,” *Phys. Med. Biol.*, vol. 62, no. 13, pp. 5556–5574, Jun. 2017.

[27] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” 2017, arXiv:1703.10593. [Online]. Available: http://arxiv.org/abs/1703.10593

**Hyoung Suk Park** received the B.S. degrees in both mathematics and biomedical engineering and the Ph.D. degree in mathematics from Yonsei University, South Korea, in 2010 and 2015, respectively. From 2015 to 2016, he was a Postdoctoral Researcher with Yonsei University. He is currently a Senior Researcher with the National Institute for Mathematical Sciences (NIMS), South Korea. His research interests are inverse problem, image processing, and deep learning for medical image analysis.

**Kyun Sang Kim** received the Ph.D. degree from the Department of Bio and Brain engineering, Korea Advanced Institute of Science and Technology (KAIST), in 2014. He was a Postdoctoral Research Fellow with the Massachusetts General Hospital (MGH) and Harvard Medical School (HMS), from 2014 to 2017, and is currently an Instructor in radiology, since 2017, at MGH and HMS. His research interests are medical image reconstruction and analysis using the compressed sensing and deep learning.

**Kiwan Jeon** received the Ph.D. degrees in mathematics from the Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 2010. Since 2011, he has been a Senior Researcher with the National Institute for Mathematical Sciences, Daejeon. His research interests are medical imaging, numerical analysis, parallel computations, and software developments.

---

**Kyoungsang Kim** received the Ph.D. degree from the Department of Bio and Brain engineering, Korea Advanced Institute of Science and Technology (KAIST), in 2014. He was a Postdoctoral Research Fellow with the Massachusetts General Hospital (MGH) and Harvard Medical School (HMS), from 2014 to 2017, and is currently an Instructor in radiology, since 2017, at MGH and HMS. His research interests are medical image reconstruction and analysis using the compressed sensing and deep learning.

**Kiwan Jeon** received the Ph.D. degrees in mathematics from the Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 2010. Since 2011, he has been a Senior Researcher with the National Institute for Mathematical Sciences, Daejeon. His research interests are medical imaging, numerical analysis, parallel computations, and software developments.