ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 978-1-7281-7605-5/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICASSP39728.2021.9413855

A BIAS-REDUCING LOSS FUNCTION FOR CT IMAGE DENOISING

Madhuri Nagare[1] *Roman Melnyk*[2] *Obaidullah Rahman*[3]
Ken D. Sauer[3] *Charles A. Bouman*[1]

[1]School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, USA 47907
[2]GE Healthcare, Waukesha, WI, USA 53188
[3]Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN, USA 46556

ABSTRACT

There is growing interest in the use of deep neural network (DNN) based image denoising to reduce patient's X-ray dosage in medical computed tomography (CT). An effective denoiser must remove noise while maintaining the texture and detail. Commonly used mean squared error (MSE) loss functions in the DNN training weight errors due to bias and variance equally. However, the error due to bias is often more egregious since it results in loss of image texture and detail. In this paper, we present a novel approach to designing a loss function that penalizes variance and bias differently. Our proposed bias-reducing loss function allows us to train a DNN denoiser so that the amount of texture and detail retained can be controlled through a user adjustable parameter. Our experiments verify that the proposed loss function enhances the texture and detail in denoised images with only a slight increase in the MSE.

***Index Terms*—** Low-dose CT, denoising, weighted mean squared error, bias reduction, deep neural networks

1. INTRODUCTION

X-ray computed tomography (CT) is perhaps the most widely used 3D medical imaging modality, and over the past decade, there has been a great deal of progress in the development of methods to further reduce noise and artifacts while improving resolution and quality [1]. In particular, noise reduction methods offer an opportunity to reduce X-ray dosage while achieving similar image quality.

A variety of methods have been studied for the reduction of noise in CT images. Iterative reconstruction algorithms provide good spatial resolution and noise reduction [2, 3]. However, iterative methods tend to be computationally expensive. An alternative approach is to apply denoising in either the sinogram [4, 5] or space domain [6, 7].

Recently, convolutional neural networks (CNNs) have become among the most popular methods for denoising CT images [8, 9]. While CNNs can be very effective, they require training with the choice of a loss function and training data. In particular, the mean squared error (MSE) loss function is commonly used for training [10] because it results in a trained network that approximately maximizes the peak signal-to-noise ratio (PSNR). However, it is also known that the MSE loss function tends to produce images that are overly smooth and lack texture [11].

One approach to improving textural details in denoised images is to use a generative adversarial network (GAN) architecture for training. The GAN attempts to produce a denoised image with the same distribution as ground truth [12, 13]. While the GANs can help to retain the texture, they could also possibly add inaccurate or even unreal image detail.

Importantly, MSE loss functions weight errors due to bias and variance equally. However, bias and variance represent very different types of errors. Bias represents systematic errors corresponding to noise or artifacts, whereas, variance is the zero mean error associated with noise. In practice, the error due to bias is often less desirable since it results in loss of image texture and detail.

In this paper, we introduce a bias-reducing loss (BR-loss) function for training DNNs that can be used to enhance texture and increase detail in the denoised images. The proposed BR-loss function reduces the weight of the variance term in the MSE, thereby producing lower bias in the denoised images. Our experimental results demonstrate that the proposed estimator learns to retain texture and recovers more structures in denoised images with only a slight increase in variance.

2. BIAS WEIGHTED MEAN SQUARED ERROR

We formulate the denoising problem in a Bayesian framework where our goal is to estimate an unknown random image X from a random noisy image Y. To do this, we will form an estimate $\hat{X} = f(Y)$ with the MSE defined by

$$\text{MSE}_{\hat{X}} = \mathbb{E}[||\hat{X} - X||^2]. \quad (1)$$

Furthermore, let $\bar{X} = \mathbb{E}[\hat{X}|X]$ be the conditional expectation of the estimate given X. Note that $\bar{X}$ is a function of ran-

This research is funded by GE Healthcare, Waukesha, WI, USA.

Authorized licensed use limited to: Purdue University. Downloaded on June 15,2022 at 13:32:15 UTC from IEEE Xplore. Restrictions apply.

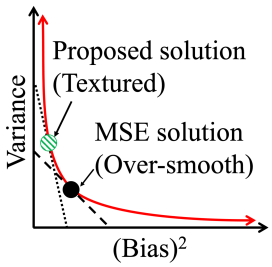

Fig. 1: Conceptual representation of optimal bias-variance trade-off for an estimator.

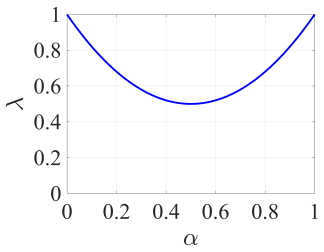

Fig. 2: λ vs. α.

dom variable X. Using the definition of $\bar{X}$, we can define the expected squared bias of the estimate as

$$\overline{\text{bias}^2_{\hat{X}}} = \mathbb{E}\left[\|\bar{X} - X\|^2\right] ,$$

and the variance of the estimate as

$$\text{variance}_{\hat{X}} = \mathbb{E}\left[\|\hat{X} - \bar{X}\|^2\right] ,$$

and then the MSE can be expressed in the following form

$$\text{MSE}_{\hat{X}} = \text{variance}_{\hat{X}} + \overline{\text{bias}^2_{\hat{X}}}. \tag{2}$$

Intuitively, the bias term is caused by systematic errors in denoising such as blurring, streaking, or other artifacts. On the other hand, the variance term represents the noisy variation in the estimate.

In practice, bias is often less desirable than variance because bias would exist even if the noisy variations were averaged out. Therefore, rather than minimizing the MSE, we propose to minimize a weighted sum of the two terms which we will refer to as the bias-weighted MSE defined as

$$\text{BW-MSE}^{\lambda}_{\hat{X}} \triangleq \lambda\, \text{variance}_{\hat{X}} + \overline{\text{bias}^2_{\hat{X}}}, \tag{3}$$

where $\lambda > 0$ specifies relative importance of the variance error. With $\lambda < 1$, we raise the relative weight of the bias error. Thus by decreasing λ, we can reduce bias in our estimate.

Fig. 1 conceptually illustrates the trade-off between the minimum bias and variance that can be achieved for an estimation problem [14]. When MSE is minimized the solution falls at the intersection of the trade-off curve with the dashed line (slope $= -1$). However, this solution might be smoother than is desirable for applications. We would prefer a solution on the optimal curve with increased variance but reduced bias, i.e., at the intersection of the trade-off curve with the dotted line (slope $= -1/\lambda$).

3. BIAS-REDUCING LOSS FUNCTION

Next we introduce a bias-reducing loss (BR-loss) function which approximates the BW-MSE. To do this, we generate

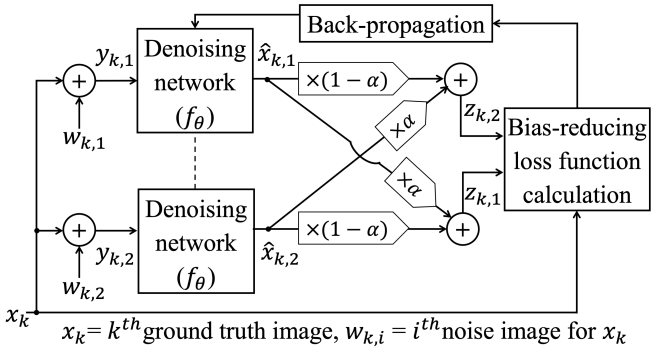

Fig. 3: Architecture for training a denoiser using the bias-reducing loss function.

a pair of noisy inputs for training by adding two independent noise realizations to the same ground truth image. More specifically, let $\{X_k\}_{k=1}^K$ be a set of i.i.d. ground truth images for training. For each ground truth image, we generate two conditionally independent noisy images, $Y_{k,1}$ and $Y_{k,2}$ with the same conditional distribution by adding two independent noise realizations to the ground truth image. Then we obtain denoised estimates as

$$\begin{aligned} \hat{X}_{k,1} &= f_\theta(Y_{k,1}) \\ \hat{X}_{k,2} &= f_\theta(Y_{k,2}) , \end{aligned}$$

where $f_\theta(\cdot)$ is a denoising algorithm with parameters θ. Note that $\hat{X}_{k,1}$ and $\hat{X}_{k,2}$ are then conditionally independent given X_k, and they have the same conditional distribution.

Then the traditional MSE loss function is given by

$$\mathcal{L}^{(MSE)}_{\theta} = \frac{1}{2K}\left\{\sum_{k=1}^{K} \|\hat{X}_{k,1} - X_k\|^2 + \|\hat{X}_{k,2} - X_k\|^2\right\} ,$$

with $\text{MSE}_{\hat{X}} = \mathbb{E}[\mathcal{L}^{(MSE)}_{\theta}]$. In order to construct the BR-loss function, we first form the following two new estimates for $\alpha \in [0, 1]$.

$$\begin{aligned} \hat{Z}_{k,1} &= \alpha\hat{X}_{k,1} + (1-\alpha)\hat{X}_{k,2}, \\ \hat{Z}_{k,2} &= (1-\alpha)\hat{X}_{k,1} + \alpha\hat{X}_{k,2}. \end{aligned}$$

Then the BR-loss function is defined as

$$\mathcal{L}^{(BR)}_{\theta,\alpha} = \frac{1}{2K}\left\{\sum_{k=1}^{K} \|\hat{Z}_{k,1} - X_k\|^2 + \|\hat{Z}_{k,2} - X_k\|^2\right\}. \tag{4}$$

Since $\hat{X}_{k,1}$ and $\hat{X}_{k,2}$ are conditionally independent given X_k,

$$Var[\hat{Z}_{k,1}|X_k] = \alpha^2 Var[\hat{X}_{k,1}|X_k] + (1-\alpha)^2 Var[\hat{X}_{k,2}|X_k],$$

where $Var[\cdot|X_k]$ is the conditional variance given X_k. Then

$$\begin{aligned} \text{variance}_{\hat{Z}} &= \mathbb{E}[\|\hat{Z}_{k,1} - \bar{Z}_{k,1}\|^2] = \mathbb{E}[Var[\hat{Z}_{k,1}|X_k]] \\ &= \left[\alpha^2 + (1-\alpha)^2\right]\mathbb{E}[Var[\hat{X}_{k,1}|X_k]] \\ &= \left[\alpha^2 + (1-\alpha)^2\right]\mathbb{E}[\|\hat{X}_{k,1} - \bar{X}_{k,1}\|^2] \\ &= \left[\alpha^2 + (1-\alpha)^2\right]\text{variance}_{\hat{X}}. \end{aligned}$$

Authorized licensed use limited to: Purdue University. Downloaded on June 15,2022 at 13:32:15 UTC from IEEE Xplore. Restrictions apply.

Furthermore,

$$\begin{aligned}\bar{Z}_{k,1} &= \mathbb{E}[\hat{Z}_{k,1}|X_k] \\ &= \alpha\mathbb{E}[\hat{X}_{k,1}|X_k] + (1-\alpha)\mathbb{E}[\hat{X}_{k,2}|X_k] = \bar{X}_{k,1}.\end{aligned}$$

Therefore, we have that $\overline{\text{bias}^2_{\hat{Z}}} = \overline{\text{bias}^2_{\hat{X}}}$. Then we have that

$$\begin{aligned}\mathbb{E}[\mathcal{L}^{(BR)}_{\theta,\alpha}] &= \text{MSE}_{\hat{Z}} = \text{variance}_{\hat{Z}} + \overline{\text{bias}^2_{\hat{Z}}} \\ &= \left[\alpha^2 + (1-\alpha)^2\right]\text{variance}_{\hat{X}} + \overline{\text{bias}^2_{\hat{X}}} \\ &= \text{BW-MSE}^{\lambda}_{\hat{X}} \ .\end{aligned}$$

So we see that the BR-loss function approximates the BW-MSE for $\lambda = \alpha^2 + (1-\alpha)^2$. Therefore, the adjustable parameter α controls the reduction in bias achieved. Fig. 2 shows the plot of λ as α varies from 0 to 1.

Fig. 3 shows the block diagram used for training with the proposed BR-loss function, where we use lower-case letters to denote realizations of the aforementioned random variables. Note that we generate two independent noise realizations for each ground truth image. Furthermore, each of the two denoising networks share the same parameters, so this is treated as a Siamese network [15] for training.

Once trained, the standalone denoising network $f_\theta(\cdot)$ is used to denoise images. So inference is done without the training architecture of Fig. 3.

4. EXPERIMENTAL RESULTS

Below we present results of a denoising CNN trained using the proposed BR-Loss function for a value of α, referred to as BR-DN-α. For comparison, we also trained the network using the conventional MSE loss, referred to as MSE-DN. To do this, we used the architecture of a CNN denoiser adopted from [16] with a single input channel and 17 convolution layers.

4.1. Methods

We acquired 29 raw clinical scans using a GE Revolution CT scanner (GE Healthcare, WI, USA), with a X-ray tube voltage and current varying from scan to scan in the range of 80-140 kVp and 40-1080 mA, respectively. The scans were reconstructed using the GE's TrueFidelity DLIR technology [17] to a slice thickness of 0.625 mm and dimension 512×512. The reconstructed volumes were used as ground truth. 20 of these volumes comprising 9776 axial slices in total were used for training and validation, while remaining 9 with 5229 slices were used for testing.

We also scanned 7 water phantoms, with a tube voltage of 120 kVp and current varying in the range of 350-380mA from scan to scan. The scans were reconstructed with filtered back-projection (FBP) to a slice thickness of 0.625mm and used to generate the noise realizations. 6 of these volumes totaling 1131 axial slices were used for training and validation, while the 7$^{\text{th}}$ volume with 249 slices was kept for testing.

In addition, we collected a low-dose clinical exam, acquired at 80 kVp tube voltage and 75 mA tube current. The exam was reconstructed with the FBP. Since low-dose scans are noisy, we used this exam for subjective evaluation.

Axial slices in the training and validation volumes were broken into 128×128 patches, with the patches randomly partitioned as 97% for training and 3% for validation. Each ground truth patch was added to two randomly selected noise patches to form two conditonally independent noisy realizations for the same ground truth patch.

To train the network, we used the Adam optimizer [18] with an initial learning rate of 0.001 and a mini-batch size of 32. The learning rate was reduced by a factor of 4 if no improvement in validation loss occurred for 5 epochs, and the training was stopped if the validation loss was not improved for 16 consecutive epochs. The network was implemented in Keras [19] and trained with two NVIDIA Tesla V100 GPUs.

Quantitative evaluation was done using the volumes kept aside for testing. The similarity of denoised images with ground truth images was quantified using the average SSIM (structural similarity) [20] and PSNR.

4.2. Discussion

Fig. 4 shows the results of applying the proposed denoising algorithm to the low-dose clinical exam. The first row of the figure shows the input noisy slices. Slices in various organs as well as in orthogonal planes have been compared to demonstrate robustness of the proposed loss function. The second and third row show denoised results with the MSE-DN and BR-DN-0.75, respectively. Notice that the proposed BR-DN-0.75 denoiser retains more texture and detail than in the MSE-DN denoised images, while still removing most of the noise.

Furthermore, the BR-DN-0.75 denoiser improves the contrast and sharpness of vessels as seen in results of the slice 1. The proposed network recovers vessels and lung fissure missing in MSE-DN results as pointed by red arrows in results for slice 2 and 3, respectively. The performance is consistent in orthogonal planes too as depicted by results of the slice 4.

Fig. 5(a) shows that the proposed BR-DN-0.75 denoising algorithm improves the average SSIM by 1.48% as compared to the MSE-DN denoiser. This is likely because the BR-DN-0.75 retains more detail and texture than the MSE-DN.

Finally, Fig. 5(b) shows the average PSNR for noisy input images, and denoised images generated with the MSE-DN and our proposed BR-DN-0.75 denoiser. Since the MSE loss function is designed to optimize the PSNR, the BR-DN-0.75 denoiser results in a PSNR that is lowered by 0.55%, which is consistent with our theoretical prediction.

Fig. 6 shows how denoised results depend on α. As the value of α is reduced from 1.0 to 0.5, the value of the variance in a uniform region of the image increases. The increase in variance along with the increase in detail is consistent with the reduction in bias. The value $\alpha = 0.75$ produced good

Authorized licensed use limited to: Purdue University. Downloaded on June 15,2022 at 13:32:15 UTC from IEEE Xplore. Restrictions apply.

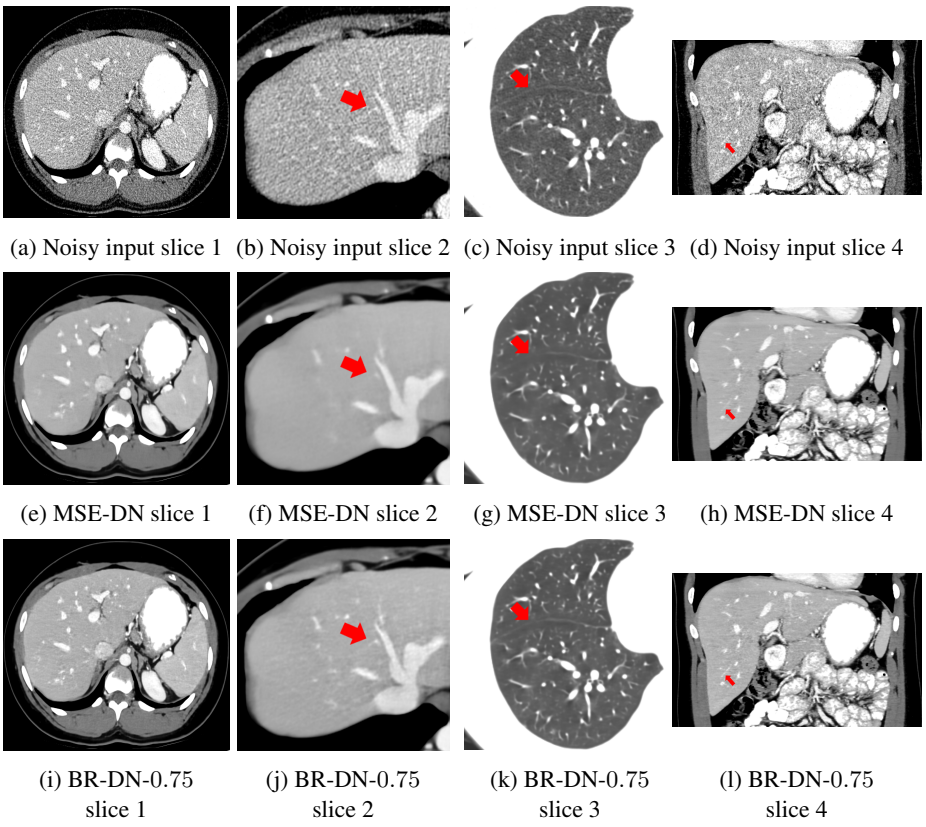

Fig. 4: Comparison of denoised results with BR-DN-0.75 and MSE-DN for the low-dose exam. Display window for slice 3 is [-700, 1000] HU and [50, 350] HU for other slices.

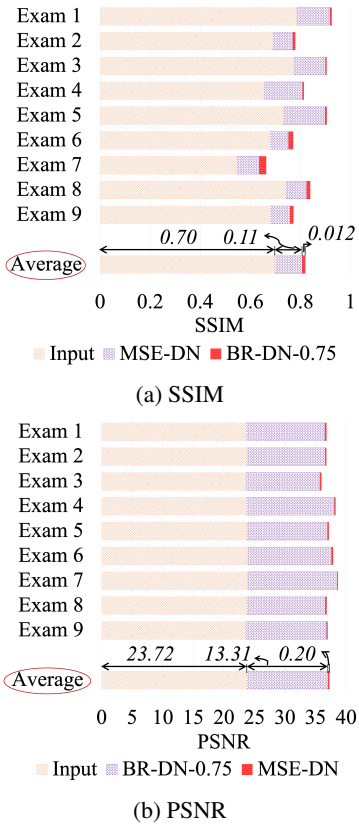

Fig. 5: Quantitative evaluation results.

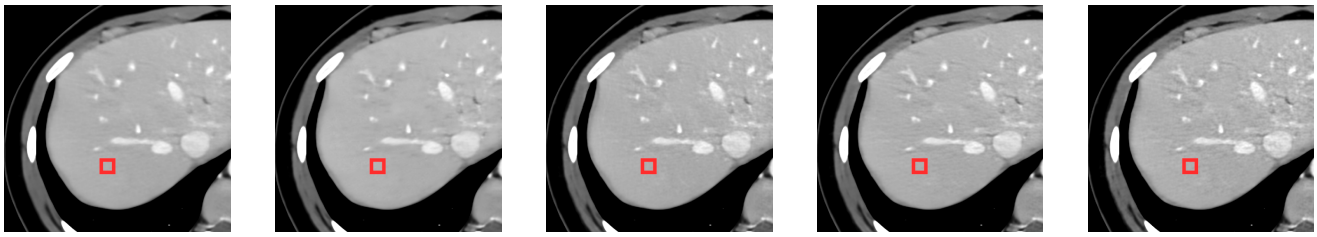

Fig. 6: Comparison of denoised results for BR-DN-α with α varying over range $[0.5, 1.0]$. Display window is [50, 350] HU.

results. However, a user can tune α for specific applications.

5. CONCLUSION

We proposed a bias-reducing loss (BR-Loss) function which weights the variance and bias terms of the mean squared error differently. The BR-Loss function works by computing a weighted average of denoised images for a pair of noisy images obtained using two independent realizations of noise, and the reduction in bias can be controlled through the choice of the weighting parameter α.

Our bias-reduced denoised images retained more texture and detail, resulting in a higher value of SSIM. However, as the theory predicted, this was at the cost of a slight decrease in the PSNR. Denoised results of a low-dose exam show that the proposed loss preserves more texture, detail, and sharpness.

6. ACKNOWLEDGMENT

We would like to thank to Jiang Hsieh, Jean-Baptiste Thibault, Nilsen Roy, Brian Nett, Jie Tang for fruitful discussions and Karen Procknow for her valuable clinical feedback.

Authorized licensed use limited to: Purdue University. Downloaded on June 15,2022 at 13:32:15 UTC from IEEE Xplore. Restrictions apply.

7. REFERENCES

[1] J. Hsieh, *Computed tomography: principles, design, artifacts, and recent advances*, SPIE press, 3^{rd} edition, 2015.

[2] J.-B. Thibault, K. Sauer, C. Bouman, and J. Hsieh, "A three-dimensional statistical approach to improved image quality for multislice helical CT," *Medical Physics*, vol. 34, no. 11, pp. 4526–4544, 2007.

[3] M. Beister, D. Kolditz, and W. Kalender, "Iterative reconstruction methods in X-ray CT," *Physica Medica*, vol. 28, no. 2, pp. 94–108, 2012.

[4] 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 Transactions on Nuclear Science*, vol. 51, no. 5, pp. 2505–2513, 2004.

[5] 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 Transactions on Medical Imaging*, vol. 25, no. 10, pp. 1272–1283, 2006.

[6] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2005, vol. 2, pp. 60–65 vol. 2.

[7] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, 2007.

[8] E. Kang, J. Min, and J. C. Ye, "A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction," *Medical Physics*, vol. 44, no. 10, pp. e360–e375, 2017.

[9] H. Chen, Y. Zhang, W. Zhang, P. Liao, K. Li, J. Zhou, and G. Wang, "Low-dose CT denoising with convolutional neural network," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, 2017, pp. 143–146.

[10] H. Chen, Y. Zhang, M. K. Kalra, F. Lin, Y. Chen, P. Liao, J. Zhou, and G. Wang, "Low-dose CT with a residual encoder-decoder convolutional neural network," *IEEE Transactions on Medical Imaging*, vol. 36, no. 12, pp. 2524–2535, 2017.

[11] J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Išgum, "Generative adversarial networks for noise reduction in low-dose CT," *IEEE Transactions on Medical Imaging*, vol. 36, no. 12, pp. 2536–2545, 2017.

[12] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M. K. Kalra, Y. Zhang, L. Sun, and G. Wang, "Low-dose CT image denoising using a generative adversarial network with wasserstein distance and perceptual loss," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1348–1357, 2018.

[13] M. Li, W. Hsu, X. Xie, J. Cong, and W. Gao, "SACNN: Self-attention convolutional neural network for low-dose CT denoising with self-supervised perceptual loss network," *IEEE Transactions on Medical Imaging*, vol. 39, no. 7, pp. 2289–2301, 2020.

[14] C. A. Bouman, *Model based image processing*, 2013.

[15] J. Bromley, J. Bentz, L. Bottou, I. Guyon, Y. LeCun, C. Moore, E. Sackinger, and R. Shah, "Signature verification using a "siamese" time delay neural network," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 7, no. 4, pp. 669–688, 1993.

[16] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2017.

[17] J. Hsieh, E. Liu, B. Nett, J. Tang, J.-B. Thibault, and S. Sahney, "A new era of image reconstruction: TrueFidelity™," Technical white paper on deep learning image reconstruction, GE Healthcare, 2019.

[18] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015*, 2015.

[19] F. Chollet et al., "Keras," `https://keras.io`, 2015.

[20] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

Authorized licensed use limited to: Purdue University. Downloaded on June 15,2022 at 13:32:15 UTC from IEEE Xplore. Restrictions apply.