Low-Dose CT via Deep CNN with Skip Connection and Network in Network

Chenyu You[a], Linfeng Yang[a], Yi Zhang[b], and Ge Wang[c]

[a]Department of Bioengineering, Stanford University, USA 94305
[b]College of Computer Science, Sichuan University, China 610065
[c]Department of Biomedical Engineering, Rensselaer Polytechnic Institute, USA 12180
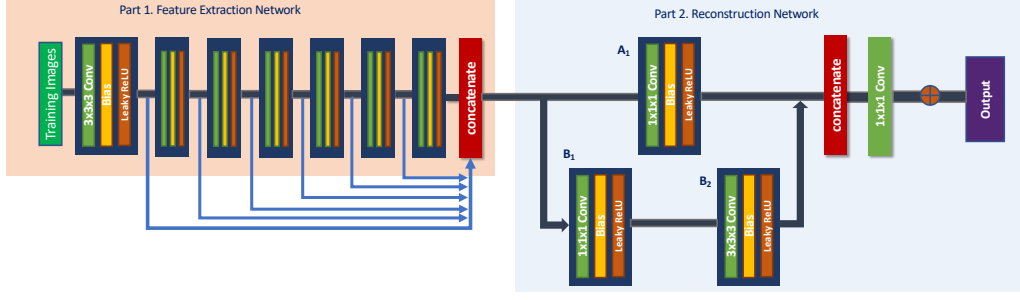
ABSTRACT

A major challenge in computed tomography (CT) is how to minimize patient radiation exposure without compromising image quality and diagnostic performance. The use of deep convolutional (Conv) neural networks for noise reduction in Low-Dose CT (LDCT) images has recently shown a great potential in this important application. In this paper, we present a highly efficient and effective neural network model for LDCT image noise reduction. Specifically, to capture local anatomical features we integrate Deep Convolutional Neural Networks (CNNs) and Skip connection layers for feature extraction. Also, we introduce parallelized 1×1 CNN, called Network in Network, to lower the dimensionality of the output from the previous layer, achieving faster computational speed at less feature loss. To optimize the performance of the network, we adopt a Wasserstein generative adversarial network (WGAN) framework. Quantitative and qualitative comparisons demonstrate that our proposed network model can produce images with lower noise and more structural details than state-of-the-art noise-reduction methods.

Keywords: Computed tomography (CT), noise reduction, deep learning, residual learning, adversarial learning.
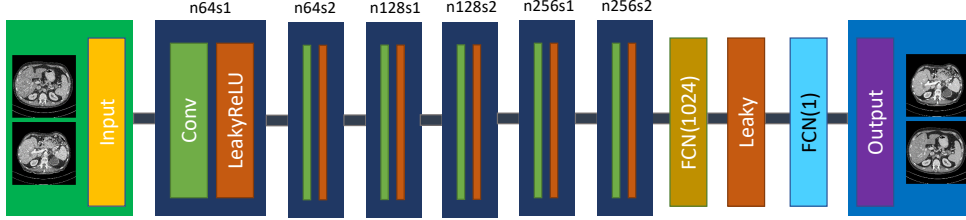
1. INTRODUCTION

X-ray computed tomography is widely used for clinical screening, diagnosis, and intervention. However, the radiation dosage associated with CT examinations may potentially induce some cancerous and genetic diseases.[1] As a result, the well-known ALARA[2] (as low as reasonably achievable) principle is universally accepted in practice, reducing unnecessary radiation exposure during medical CT imaging. One of the commonly-used methods is to lower the X-ray flux towards the x-ray detector array by adjusting the milliampere-seconds (mAs) and kVp settings for data acquisition. However, since CT imaging is a quantum integration process, an insufficient amount of photons will introduce excessive statistical noise and significantly deteriorate image quality. Therefore, how to preserve image quality for clinical tasks at minimum radiation dose has been one of the major endeavors in the CT field over the past decade.

Deep learning (DL) has been now applied in almost all medical tomographic imaging areas, inspired by a large amount of image processing results.[3,4] In particular, several DL-based studies for image noise reduction were performed.[5–9] Since CNN models learn high-level representations in terms of multiple layers of feature abstraction from big training images, it is expected to have a better denoising capability than other classic image-domain methods. In this paper, we aim to maintain anatomical and pathological information, and at the same time suppress image noise due to low radiation dose. Specially, we develop a new ConvNet architecture for LDCT denoising. In order to progressively capture both local and global anatomical features, here we design cascaded subnetworks to integrate complementary textural information. Moreover, by introducing residual learning at the image reconstruction stage, the network model is made to learn the residuals between a bicubic interpolation image and the corresponding full-dose CT (FDCT) image so that the denoising performance can be boosted. Finally, with parallelized CNNs (Network in Network) local patches within the receptive field are effectively analyzed.[10] As far as the loss function is concerned, we introduce the L_1 norm instead of L_2 distance to disencourage blurring.[11]

(a)Architecture of the generator G

(b)Architecture of the discriminator D

Figure 1. Our proposed network structure. The G is composed of a feature extraction network and a reconstruction network. Note that the reconstruction block A_1 is denoted as Channel A, the reconstruction blocks B_1 and B_2 as Channel B, n stands for the number of convolutional kernels, and s for convolutional stride. For example, $n32s1$ means that the convolutional layer has 32 kernels with stride 1.

2. METHODS

Let a vector $\boldsymbol{x} \in \mathbb{R}^{M \times 1}$ represent a noisy LDCT image of $N \times N$ pixels, and a vector $\boldsymbol{y} \in \mathbb{R}^{M \times 1}$ its corresponding NDCT image, $M = N \times N$. A DL-based network model $\boldsymbol{DL}$ with the multiple processing layers is trained to process LDCT images according to a non-linear input-output mapping, which is equivalent to solve the following optimization problem:

$$\arg\min_{\boldsymbol{DL}} ||\boldsymbol{DL}(\boldsymbol{x}) - \boldsymbol{y}||_1 \tag{1}$$

Our network constitutes two components: the generative model G and the discriminative model D as shown in Fig. 1. In the feature extraction network, the number of filters are $32, 26, 22, 18, 14, 11, 8$ for the Conv layers respectively. Also, in the image reconstruction network, the three channels are cascaded. The reconstruction block A_1, B_1, B_2 consist of $24, 8, 8$ filters respectively. Because all the outputs from the feature extraction layers were densely connected, and the final outputs after reconstruction is large, therefore we introduce 1×1 CNN after the reconstruction network to reduce the input dimension and decrease computational complexity. Instead of constructing a high-quality image by the network itself, we incorporate residual learning strategy to capture high-frequency features that can help improve the quality of low-dose CT images.[12]

The covariance of pixel level features will significantly influence the denoising performance.[11] Indeed, in our experiments the pure CNN-based model tends to produce blurry features. GANs *et al.*[13] is a promising approach to address the aforementioned limitations, since GAN is a framework for generative modeling of data through minimizing the discrepancy between the prior data distribution P_z of the generated outputs from G and the real data distribution P_r. Hence, we force the denoised image to stay on the image manifold by matching the distribution of real images to that of synthesized input images. Even though GAN has been widely applied in image processing, they suffer from model divergence and are unstable to train.[14] To regularize the training process for GAN, we adopt the Earth Moving distance (EM distance), instead of the original Jensen-Shannon
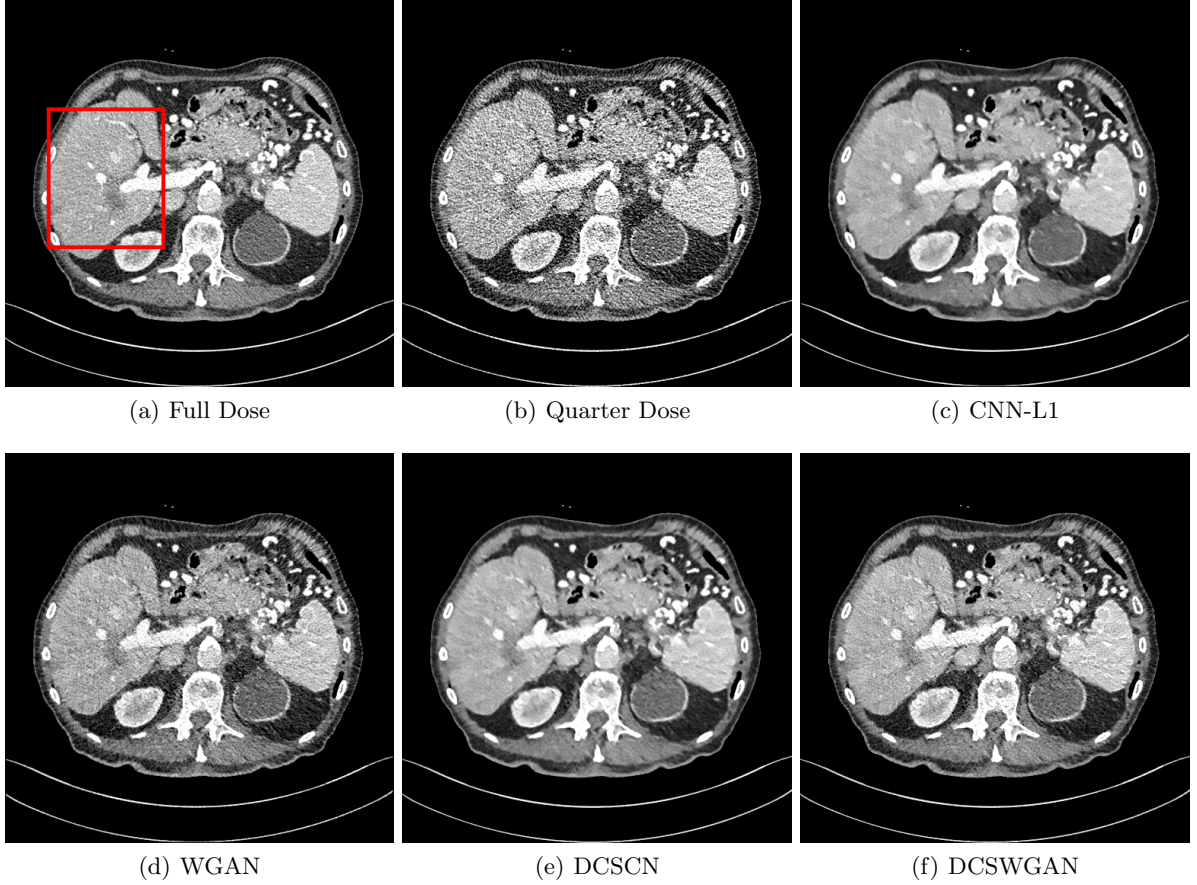
Figure 2. Results with abdomen CT images.(a) FDCT, (b) LDCT, (c) CNN-L1, (d) WGAN, (e) DCSCN, and (h) DCSWGAN. The red box indicates the region zoomed in Fig. 3. This display window is [-160, 240]HU.
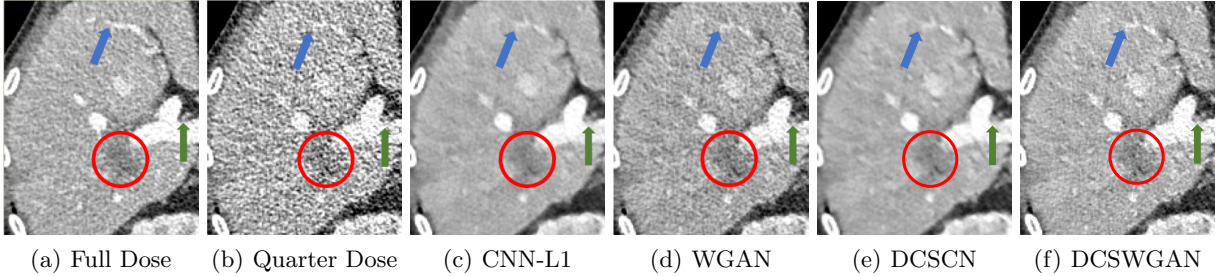

Figure 3. Zoomed regions of interest(ROIs) marked by the red box in Fig. 2. (a) FDCT, (b) LDCT, (c) CNN-L1, (d) WGAN, (e) DCSCN, and (f) DCSWGAN. The dashed circle shows the metastasis, and the green and blue arrows indicate two subtle structural features. The display window is [-160,240]HU

(JS) divergence, in the objective function.[15] Thus, the adversarial loss is formulated as:

$$\min_G \max_D L_{\text{WGAN}}(D, G) = -\mathbb{E}_{\boldsymbol{y}}[D(\boldsymbol{y})] + \mathbb{E}_{\boldsymbol{x}}[D(G(\boldsymbol{x}))] + \lambda \mathbb{E}_{\hat{\boldsymbol{x}}}[(||\nabla_{\hat{\boldsymbol{y}}} D(\hat{\boldsymbol{y}})||_2 - 1)^2], \tag{2}$$

where the first two terms are for the Wasserstein estimation, the third term penalizes the deviation of the gradient norm with respect to the input from one, $\hat{\boldsymbol{y}}$ is uniformly sampled along straight lines pairs of denoised and real images, and λ is a regularization parameter.

Although L_1 and L_2 losses are both the mean-based loss function, the effects of these two loss functions differ in terms of denoising. Compared with the L_2 loss, the L_1 loss neither over-penalize large differences nor tolerate
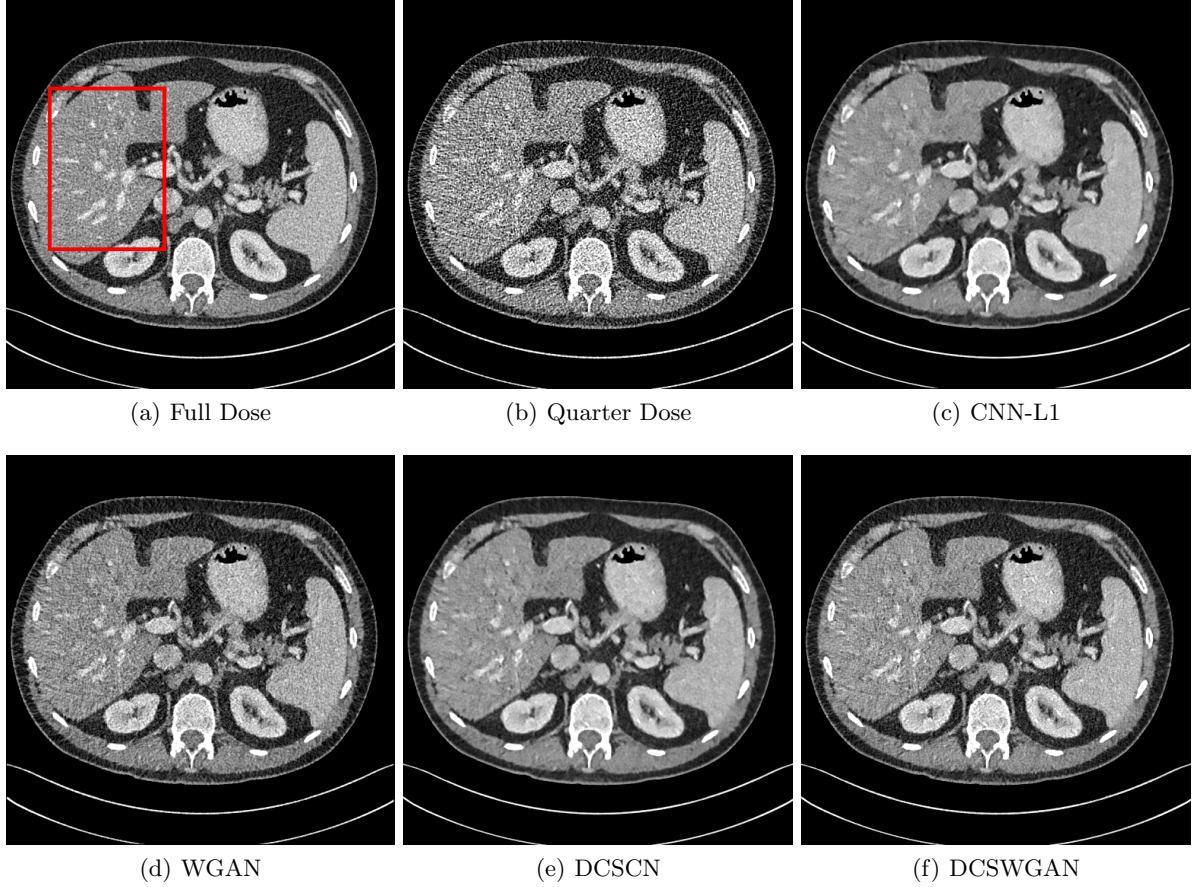
Figure 4. Results from abdomen CT images.(a) FDCT, (b) LDCT, (c) CNN-L1, (d) WGAN, (e) DCSCM, and (f) DCSWGAN. The red box indicates the region of interest zoomed in Fig. 5. This display window is [-160, 240]HU.
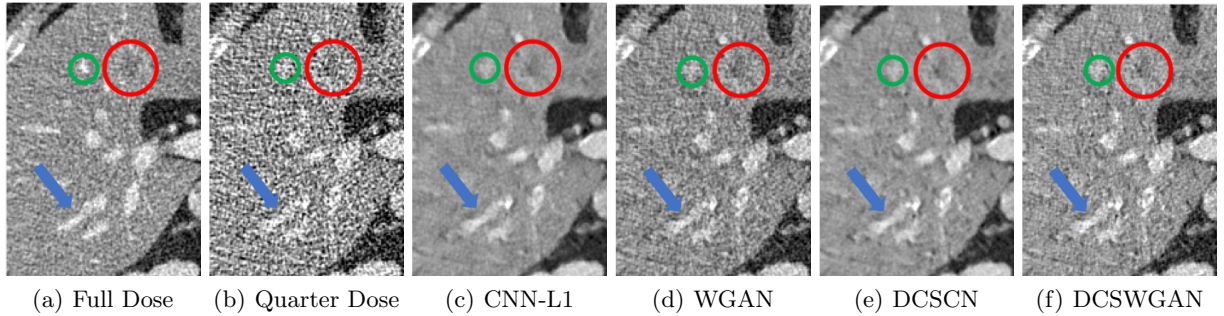

Figure 5. Zoomed regions of interest (ROIs) marked by the red box in Fig. 4. (a) FDCT, (b) LDCT, (c) CNN-L1, (d) WGAN, (e) DCSCM, and (f) DCSWGAN. The dashed circle indicates the metastasis, and the green and blue arrows show two subtle structural features. The display window is [-160,240]HU

small errors between denoised images and the gold-standard. Thus, the L_1 loss alleviates some limitations of the L_2 loss. Additionally, the L_1 loss shares the same merits that the L_2 loss has; e.g, a fast convergence speed.

The L_1 loss is formulated as follows:

$$L_1(G) = \frac{1}{HWD} \mid G(\boldsymbol{x}) - \boldsymbol{y} \mid \qquad (3)$$

where H, W, D stand for the height, width, and depth of a 3D image patch, respectively, $\boldsymbol{y}$ denotes a gold-

Table 1. Quantitative results associated with different approaches in Figs. 2 and 4.

	Fig. 2			Fig. 4		
	PSNR	SSIM	RMSE	PSNR	SSIM	RMSE
LDCT	22.818	0.761	0.0723	21.558	0.659	0.0836
CNN-L1	27.791	0.822	0.0408	26.794	0.738	**0.0457**
WGAN	25.727	0.801	0.0517	24.655	0.711	0.0585
DCSCN	**28.016**	**0.883**	**0.0397**	**26.943**	0.730	0.0530
DCSWGAN	26.928	0.828	0.0449	25.721	0.808	0.0517

standard image (NDCT), and $G(\boldsymbol{x})$ represents a denoised image from a LDCT image $\boldsymbol{x}$.

Besides, there are two aspects in the sparse representation step for image denoising, which are the prior information level and the sparsity level. We first introduce the adversarial loss to capture local anatomical information. Then, we use L_1 loss to improve the sparsity of our representation, leading to the solution of the following optimization problem.

Leveraging Eqs. (2) and (3) together, the overall joint objective function is formulated as:

$$L_{\text{obj}} = \min_G \max_D L_{\text{WGAN}}(D, G) + \lambda_1 L_1(G) \tag{4}$$

where λ_1 is a regularization parameter to balance the information preservation and the sparsity-promotion between the WGAN adversarial loss and the L_1 loss.

3. EXPERIMENTAL RESULTS

To evaluate the effectiveness of the proposed method, we compared it with existing state-of-the-art denoising methods, including CNN-L1 (L_1-net)[11] and WGAN-based CNN.[6] Note that all the parameters of these selected benchmark methods were set to that suggested in the original papers. For brevity, we denote our Deep CNN with Skip Connection and Network in Network as DCSCN, and the model using a Wasserstein Generative Adversarial Network as DCSWGAN.

The experiment set-up is as follows. First, to minimize the generalization error, we adopted leave-one-out cross-validation to refine the denoising performance. Then, in the training phase, $499,996$ pairs of image patches of size 80×80 from 7 patients were randomly selected. For validation, $5,096$ pairs of image patches were extracted from other 3 patients and set to the same size. It is worth noting that the size of extracted patches was made large enough to include regions of liver lesions. Next, in addition to preserve the integrity of data, here we scaled the CT Hounsfield Value (HU) to the unit interval [0,1] before the images were fed to the network. Finally, we used three common image quality metrics: peak signal-to-noise ratio (PSNR), structural similarity index (SSIM),[16] and root-mean-square error (RMSE) to evaluate the denoised image quality.

The visual inspection of our results indicates that the LDCT images in Figs. 2(b) and 4(b) have strong background noises. Furthermore, we find that the l_1-net has a great noise suppression capability, but it still has over-smoothing effects on some textural details in the ROIs in Fig. 3(c). The l_1-net achieved a high signal-to-noise ratio (SNR), but it yielded lower contrast resolution. From ROIs in Fig. 5(c), it is seen that there are still some blocky effects marked by the blue arrow. Figs. 2(d) and 4(d) display the WGAN-processed denoised LDCT images with improving structural identification. However, as shown in Figs. 3(d) and 5(d), the WGAN model also introduced strong image noise. In Figs. 2(e) and 4(e), the proposed DCSCN achieved noise reduction but also suffered from image blurring . As shown in Fig. 2(f) and 4(f), our proposed DCSWGAN network model demonstrates the best performance in noise reduction and feature preservation as compared to all the competing denoisng methods. Figs. 3(f) and 5(f) illustrate that DCSWGAN not only effectively suppressed strong noise but also kept subtle textural information, outperforming other denoising models; see ROIs (in Figs. 3 and 5) and/or zoom in for better visualization.

The PSNRs, SSIMs, and RMSEs are listed in Table 1. For noise reduction, the performance metrics were significantly improved by our proposed method (DCSCN). This demonstrates that using residual learning steak

artifacts and image noise can be largely removed, enhancing the image quality. In this pilot study, DCSCN achieved the best performance in terms of PSNR and SSIM, and preserved anatomical features the most faithfully. However, there still exits blurry effects as shown in Figs. 3 and 5. DCSWGAN obtained the second best results in term of SSIM. It is noted that our method DCSWGAN produced visually pleasant results with sharp edges.

4. CONCLUSION

In this work, we have proposed a CNN-based network with skip-connection and network in network to capture structural information and suppress image noise. First, both local and global features are cascaded through skip connections before passing to the reconstruction network. Then, multi-channels are introduced for the reconstruction network with different local receptive fields to optimize the reconstruction performance. Also, the network in network technique is applied to lower the computational complexity. Our results have suggested that the proposed method could be generalized to various medical image denoising problems but further efforts are needed for training, validation, testing, and optimization.

REFERENCES

[1] de González, A. B., Mahesh, M., Kim, K.-P., Bhargavan, M., Lewis, R., Mettler, F., and Land, C., "Projected cancer risks from computed tomographic scans performed in the united states in 2007," *Arch. Intern. Med.* **169**(22), 2071–2077 (2009).

[2] Brenner, D. J. and Hall, E. J., "Computed tomography - an increasing source of radiation exposure," *New Eng. J. Med.* **357**(22), 2277–2284 (2007).

[3] Wang, G., "A perspective on deep imaging," *IEEE Access* **4**, 8914–8924 (2016).

[4] Wang, G., Kalra, M., and Orton, C. G., "Machine learning will transform radiology significantly within the next 5 years," *Med. Phys.* **44**(6), 2041–2044 (2017).

[5] Wolterink, J. M., Leiner, T., Viergever, M. A., and Išgum, I., "Generative adversarial networks for noise reduction in low-dose CT," *IEEE Trans. Med. Imaging* **36**(12), 2536–2545 (2017).

[6] Yang, Q., Yan, P., Zhang, Y., Yu, H., Shi, Y., Mou, X., Kalra, M. K., Zhang, Y., Sun, L., and Wang, G., "Low-dose CT image denoising using a generative adversarial network with wasserstein distance and perceptual loss," *IEEE Trans. Med. Imaging* **37**(6), 1348–1357 (2018).

[7] Kang, E., Min, J., and Ye, J. C., "A deep convolutional neural network using directional wavelets for low-dose x-ray ct reconstruction," *Med. Phys.* **44**(10), e360–e375 (2017).

[8] You, C., Li, G., Zhang, Y., Zhang, X., Shan, H., Li, M., Ju, S., Zhao, Z., Zhang, Z., Cong, W., et al., "CT Super-resolution GAN Constrained by the Identical, Residual, and Cycle Learning Ensemble (GAN-CIRCLE)," *IEEE Trans. Med. Imaging* (2019).

[9] Lyu, Q., You, C., Shan, H., and Wang, G., "Super-resolution MRI through Deep Learning," *arXiv preprint arXiv:1810.06776* (2018).

[10] Lin, M., Chen, Q., and Yan, S., "Network in network," *Int. Conf. Learn. Representations. (ICLR)* (2014).

[11] You, C., Yang, Q., Shan, H., Gjesteby, L., Guang, L., Ju, S., Zhang, Z., Zhao, Z., Zhang, Y., Cong, W., and Wang, G., "Structure-sensitive multi-scale deep neural network for low-dose CT denoising," *IEEE Access* (2018).

[12] Yu, H., Liu, D., Shi, H., Yu, H., Wang, Z., Wang, X., Cross, B., Bramler, M., and Huang, T. S., "Computed tomography super-resolution using convolutional neural networks," in [*Proc. IEEE Intl. Conf. Image Process.*], 3944–3948 (2017).

[13] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y., "Generative adversarial nets," in [*Proc. Adv. Neural Inf. Process. Syst.*], 2672–2680 (2014).

[14] Radford, A., Metz, L., and Chintala, S., "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434* (2015).

[15] Arjovsky, M., Chintala, S., and Bottou, L., "Wasserstein GAN," *arXiv preprint arXiv:1701.07875* (2017).

[16] Wang, Z., Simoncelli, E. P., and Bovik, A. C., "Multiscale structural similarity for image quality assessment," in [*Proc. IEEE Asilomar Conf. Signals, Syst., Comput.*], **2**, 1398–1402, Ieee (2003).