

The SunCHECK™ Platform powers Quality Management in radiation therapy.

Scalable to meet the needs of any clinic or network, SunCHECK helps:

-  **Reduce risks**
-  **Control costs**
-  **Improve treatment quality**

Visit sunnuclear.com to learn how SunCHECK fits into your clinic.

A convolutional neural network for ultra-low-dose CT denoising and emphysema screening

Tingting Zhao

Department of Radiation Oncology, University of California, Los Angeles, CA 90095, USA

Michael McNitt-Gray

Departments of Biomedical Physics and Radiology, University of California, Los Angeles, CA 90095, USA

Dan Ruan^{a)}

Department of Radiation Oncology, University of California, Los Angeles, CA 90095, USA

(Received 5 March 2019; revised 3 May 2019; accepted for publication 21 May 2019; published 17 July 2019)

Purpose: Reducing dose level to achieve ALARA is an important task in diagnostic and therapeutic applications of computed tomography (CT) imaging. Effective image quality enhancement strategies are crucial to compensate for the degradation caused by dose reduction. In the past few years, deep learning approaches have demonstrated promising denoising performance on natural/synthetic images. This study tailors a neural network model for (ultra-)low-dose CT denoising, and assesses its performance in enhancing CT image quality and emphysema quantification.

Methods: The noise statistics in low-dose CT images has its unique characteristics and differs from that used in general denoising models. In this study, we first simulate the paired ultra-low-dose and targeted high-quality image of reference, with a well-validated pipeline. These paired images are used to train a denoising convolutional neural network (DnCNN) with residual mapping. The performance of the DnCNN tailored to CT denoising (DnCNN-CT) is assessed over various dose reduction levels, with respect to both image quality and emphysema scoring quantification. The possible over-smoothing behavior of DnCNN and its impact on different subcohort of patients are also investigated.

Results: Performance evaluation results showed that DnCNN-CT provided significant image quality enhancement, especially for very-low-dose level. With DnCNN-CT denoising on 3%-dose cases, the peak signal-to-noise ratio improved by 8 dB and the structure similarity index increased by 0.15. This outperformed the original DnCNN and the state-of-the-art nonlocal-mean-type denoising scheme. Emphysema mask was also investigated, where lung voxels of abnormally low attenuation coefficient were marked as potential emphysema. Emphysema mask generated after DnCNN-CT denoising on 3%-dose image was demonstrated to agree well with that from the full-dose reference. Despite over-smoothing in DnCNN denoising, which contributed to slight underestimation of emphysema score compared to the reference, such minor overcorrection did not affect clinical conclusions. The proposed method provided effective detection for cases with appreciable emphysema while serving as a reasonable correction for normal cases without emphysema.

Conclusions: This work provides a tailored DnCNN for (ultra-)low-dose CT denoising, and demonstrates significant improvement on both the image quality and the clinical emphysema quantification accuracy over various dose levels. The clinical conclusion of emphysema obtained from the denoised low-dose images agrees well with that from the full-dose ones. © 2019 American Association of Physicists in Medicine [<https://doi.org/10.1002/mp.13666>]

Key words: deep network, emphysema screening, low-dose CT, quantitative imaging

1. INTRODUCTION

Computed tomography (CT), an important platform for quantitative imaging, is associated with risk of ionizing radiation, especially with repetitive screening scans. The use of low-dose CT is recommended in medical societies for regular screening examinations of high-risk population to reduce the lung cancer mortality.¹ The image quality degradation caused by decreased mAs acquisitions, however, may reduce its value in disease assessment and screening studies.^{2,3} Therefore, it is highly desirable to achieve high-quality images with reduced dose, using advanced noise reduction development.

Compared to sinogram-domain approaches, denoising methods working on reconstructed image data have the advantage of being more convenient and practical as proprietary raw sinogram data is usually either inaccessible or challenging to obtain.

Image denoising has been a long-standing image processing topic. For natural and artificial images, the state-of-the-art techniques include model-driven approaches with embedded sparsity^{6,7} or nonlocal approaches such as the renowned block-matching and three-dimensional (3D) filtering (BM3D) to take advantage of self-similarity.^{4,5} During the last couple of years, deep learning networks have

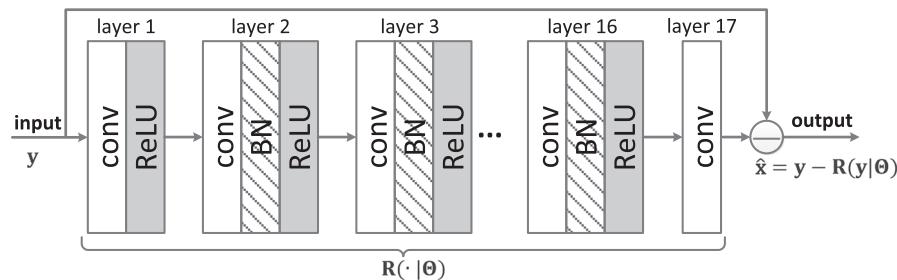


FIG. 1. The structure of DnCNN with residual mapping. The key modules include convolutional filtering (conv), ReLU, batch normalization (BN), and residual mapping. The first 16 convolutional layers have 64 filters with kernel size of 3×3 and stride of 1. The last convolutional layer has only one filter with the same setup.

presented themselves as a promising alternative to explicitly model-driven methods. The earliest studies have used convolutional neural network (CNN) in natural image denoising, and achieved comparable performance to the BM3D method in the presence of Gaussian noise, but was shown to be inferior on more realistic and complicated noise.^{8–10} The incorporation of residual learning and batch normalization gave rise to an improved denoising network, known as DnCNN, yielding better results for blind Gaussian denoising tasks with unknown noise statistics.^{11,12}

In this study, we apply the general DnCNN method to the challenging task of ultra-low-dose CT denoising, where the noise is object-dependent and complex artifacts such as streak exist.¹³ We use a verified pipeline to generate low-dose counterparts for each CT image of reference quality and tailor a DnCNN based on such paired image set.¹⁴ In this paper, we use DnCNN-CT to indicate the tailored DnCNN for the specific purpose of recovering high-quality “equivalent” CT scans from low-dose acquisitions. We compare the quality of the denoised images using DnCNN-CT with that using the original DnCNN and the state-of-the-art BM3D denoising, over different dose levels.^{4,12} We further investigate their clinical effect on low-dose thoracic CT screening for emphysema, where dose level reduction severely degrades the recognition of lung structures and quantitative analysis of airway/emphysema measures.^{15,16} We visualize the potential emphysema detected with abnormally low attenuation coefficients, and compare the clinically used diagnostic measures, including (a) relative area (RA) of lung with the intensity lower than a specified threshold in Hounsfield unit (HU), (b) percentile (PERC) as the cutoff HU value below which a specified percentage of voxels lie, and (c) mean lung intensity.^{17,18} We also assess the divergence of the lung intensity distribution from the full-dose reference. In addition, statistical hypothesis test is performed on these quantitative emphysema measures to determine the difference between the DnCNN-CT and other denoising candidates.

The rest of this paper is organized as follows. Section 2 briefly introduces the structure of the denoising network. Section 3 reports experiment results. Section 4 summarizes this work.

2. METHODS

2.A. Network structure of DnCNN

We start from a DnCNN model for natural image denoising, which has been pretrained on 400 gray images of size 180×180 and Gaussian noise level of 25.¹² In this work, this original DnCNN is tailored based on simulated CT scans. CT scans are assumed to be disturbed by additive low-dose noise, denoted by $\mathbf{y} = \mathbf{x} + \mathbf{n}$, where \mathbf{y} is the low-dose image, \mathbf{x} is the component of full-dose image, and \mathbf{n} is the noise.

DnCNN benefits from residual learning and batch normalization, with learning speedup and performance enhancement in handling general image denoising problem. Instead of directly seeking an inversion model $H(\mathbf{y}) \approx \mathbf{x}$, DnCNN learns a residual mapping $R(\mathbf{y}) \approx \hat{\mathbf{x}}$ and generates the denoised image via $\hat{\mathbf{x}} = \mathbf{y} - R(\mathbf{y})$. This design effectively addresses the vanishing gradient problem in very deep networks, yielding more efficient training and more accurate results.¹¹

Figure 1 shows the framework of DnCNN. The input patch size in training is 40×40 in this study. The DnCNN has 17 layers, which is an empirical choice following the work in Ref.¹² For the first 16 layers, 64 convolutional filters are used with size of 3×3 , and rectified linear unit (ReLU) is used for nonlinearity. Batch normalization is performed between

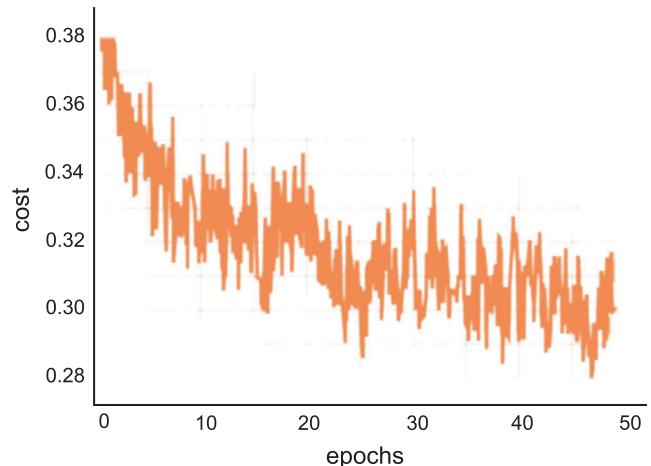


FIG. 2. The change of cost during the training of DnCNN-CT. [Color figure can be viewed at wileyonlinelibrary.com]

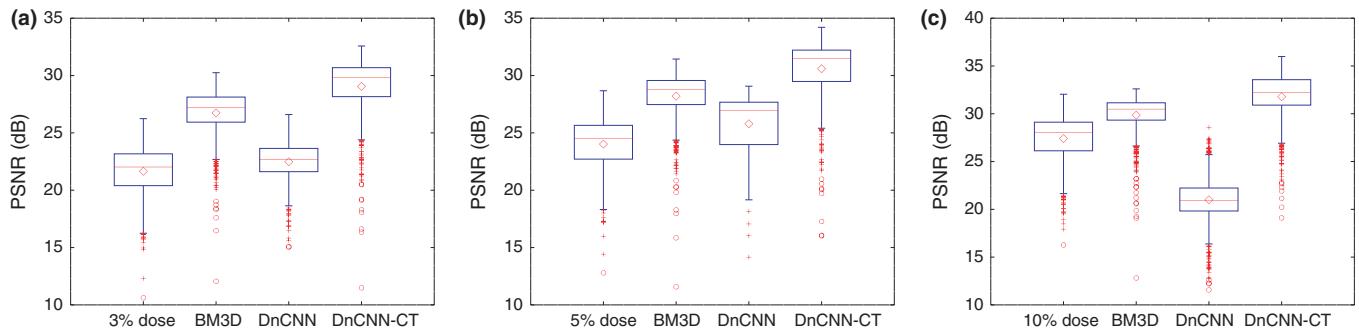


Fig. 3. Boxplots of PSNR for (a) 3%-dose, (b) 5%-dose, and (c) 10%-dose cases, compared among the input image, the estimate after BM3D denoising and that after the DnCNN(-CT). [Color figure can be viewed at [wileyonlinelibrary.com](#)]

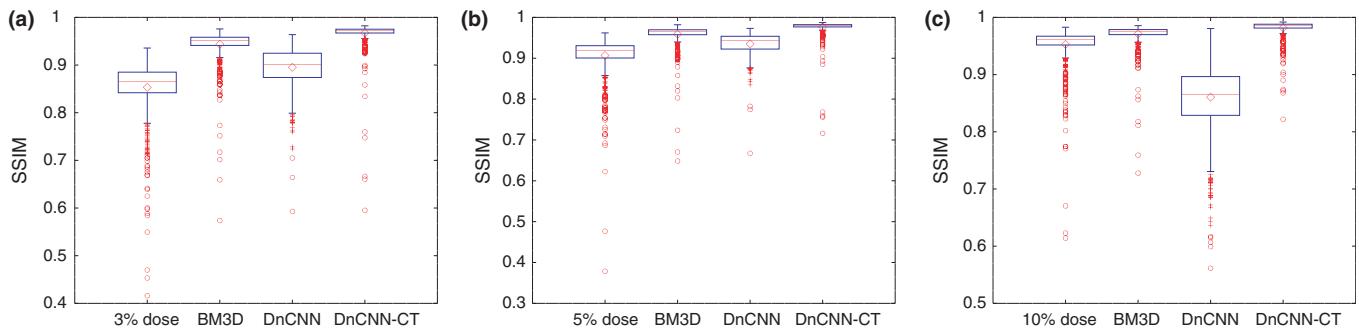


Fig. 4. Boxplots of SSIM for (a) 3%-dose, (b) 5%-dose, and (c) 10%-dose cases, compared among the input image, the estimate after BM3D denoising and that after the DnCNN(-CT). [Color figure can be viewed at [wileyonlinelibrary.com](#)]

the convolution unit and ReLU in all the middle layers to reduce the internal covariate shift, which brings benefit in training speedup and performance improvement.¹⁹ The cost function is formulated as the accumulated mean squared error compared to the true noise \mathbf{n} , that is,

$$\Theta = \arg \min_{\Theta} \sum_{i=1}^N \|R(\mathbf{y}_i|\Theta) - (\mathbf{y}_i - \mathbf{x}_i)\|^2,$$

where Θ is the set of network parameters, N is the batch size, and $(\mathbf{x}_i, \mathbf{y}_i)$ is the i th pair of full-dose patch and the corresponding low-dose version.

2.B. Materials

The lung CT scans used in this study were collected from a multidetector CT scanner (Definition AS, Siemens Healthineers, Forchheim Germany) with original dose of 2 mGy. Images had been reconstructed at thickness of 1.0 mm using the B45 filter from the *FreeCT_wFBP* software package.¹⁴ The reduced-dose scans were obtained by adding calibrated noise to the raw projection data, based on the algorithm proposed in²⁰ and the noise model validated in.²¹ For each case, volumetric image datasets with both the original (2 mGy) dose and the simulated dose reduced to 3% (0.06 mGy), 5% (0.1 mGy), 10% (0.2 mGy), 25% (0.5 mGy), 50% (1.0 mGy), and 75% (1.5 mGy) of original, were provided. The dataset used in this study contained six cases, among which three cases had pronounced emphysema. We arbitrarily chose three cases (including cases with and

without emphysema) for training, from which 30% of the patches were randomly selected as validation data during the training stage. The remaining three cases to evaluate the denoising performance. Each case had around 200 slices with slice size of 512×512 , and the denoising was performed in two-dimensions (2D).

Mini-batch stochastic gradient descent (SGD) was used to train the original DnCNN model with a batch size of 128. The network was initialized with the weights of the original DnCNN,¹² and trained in 50 epochs with a weight decay of 1e-04, momentum of 0.9, and fixed learning rate of 1e-04. The change of cost function in 50 epochs during the training stage is illustrated in Fig. 2.

We compared the performance of DnCNN-CT with (a) the state-of-the-art non-local-mean-type BM3D filtering scheme and (b) the original DnCNN. The peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) index were calculated by comparing a denoised low-dose image $\hat{\mathbf{x}}$ with the original full-dose one \mathbf{x} .

$$\text{PSNR}(\hat{\mathbf{x}}, \mathbf{x}) = \frac{\max_{z \in \Omega} |x_z|^2}{\frac{1}{|\Omega|} \int_{z \in \Omega} |\hat{x}_z - x_z|^2 dz} \quad (1)$$

$$\text{SSIM}(\hat{\mathbf{x}}, \mathbf{x}) = \frac{(2\mu_{\hat{x}}\mu_x + c_1)(2\sigma_{\hat{x}x} + c_2)}{(\mu_{\hat{x}}^2 + \mu_x^2 + c_1)(\sigma_x^2 + \sigma_{\hat{x}}^2 + c_2)} \quad (2)$$

We evaluated both PSNR and SSIM within the lung region, denoted by Ω , which was the region of interest in the subsequent emphysema scoring task. The μ_x (σ_x^2) and $\mu_{\hat{x}}$ ($\sigma_{\hat{x}}^2$) were the mean (variance) of lung intensity value from the full-dose

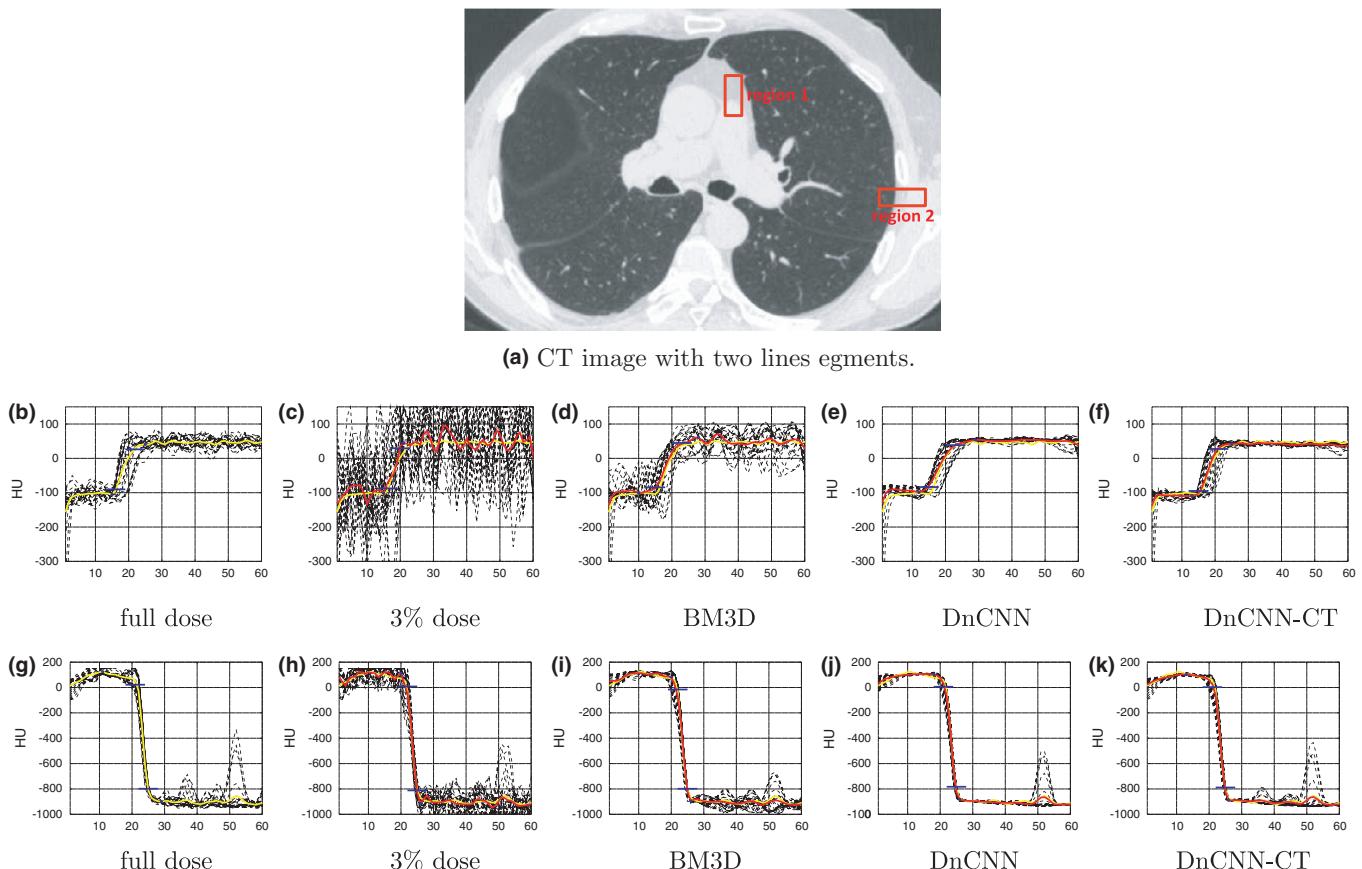


FIG. 5. Edge profiles for two illustrated boundaries, as annotated in (a). Region 1: (b)-(f), region 2: (g)-(k). Edge profiles along the lines across the boundary are shown in black dashed line, with their moving average shown in red and full dose reference in yellow. The distance for the averaged edge response to change from 10% to 90%, shown as the horizontal blue lines, is coarsely measured as 7, 5, 10, 9, 7 voxels in region #1 and 5, 7, 5, 5, 5 in region #2 for the full-dose image, the low-dose image, and the denoised images by BM3D, the original DnCNN and DnCNN-CT, respectively. [Color figure can be viewed at wileyonlinelibrary.com]

and the denoised low-dose image, respectively, and $\sigma_{\hat{x}x}$ was the covariance between them. The $c_1 = (0.01L)^2$ and $c_2 = (0.03L)^2$ were regularization constants with L being the value range of lung intensity.* We also characterized image resolution change after denoising by the edge response in regions with boundaries, across which image contrast was different.

For the subsequent emphysema scoring task, four quantitative measures based on lung intensity were evaluated/compared^{17,18}:

1. RA is defined as the relative lung area falling below an HU threshold.^{22,23} In this paper, we evaluated RA970, RA950, RA930, and RA910, indicating the RA with lung intensity below -970, -950, -930, and -910 HU, respectively.
2. PERC indicates the cutoff HU value below which a specified percentage of voxels within lungs are

distributed. In this study, PERC10, PERC15, and PERC20 for 10%, 15%, and 20% percentiles were evaluated, respectively.

3. Mean lung intensity in HU was also collected and compared. For cases with emphysema, the distribution of lung intensity shifts towards the low-value region, resulting in smaller mean lung intensity as well as higher RA and smaller PERC results.
4. Kullback–Leibler (KL) divergence was also reported as a full assessment of distributional agreement between the lung intensity of (denoised) low-dose images and that of full-dose reference

$$D_{\text{KL}}(P_{\hat{x}} || P_x) = \int p_{\hat{x}}(z) \log \frac{p_{\hat{x}}(z)}{p_x(z)} dz, \quad (3)$$

where $P_{\hat{x}}$ and P_x were the probability distributions of the lung intensity value from the (denoised) low-dose image and that from the full-dose one, respectively.

In additional, statistical hypothesis test was also performed on measures above to show the difference between DnCNN-CT and other denoising approaches.

*The image quality evaluation in this study was performed on lung intensity in HU. The value of L was defined as the difference between the maximum and the minimum intensity value, in the range of [200, 1200] HU for the dataset used in this study.

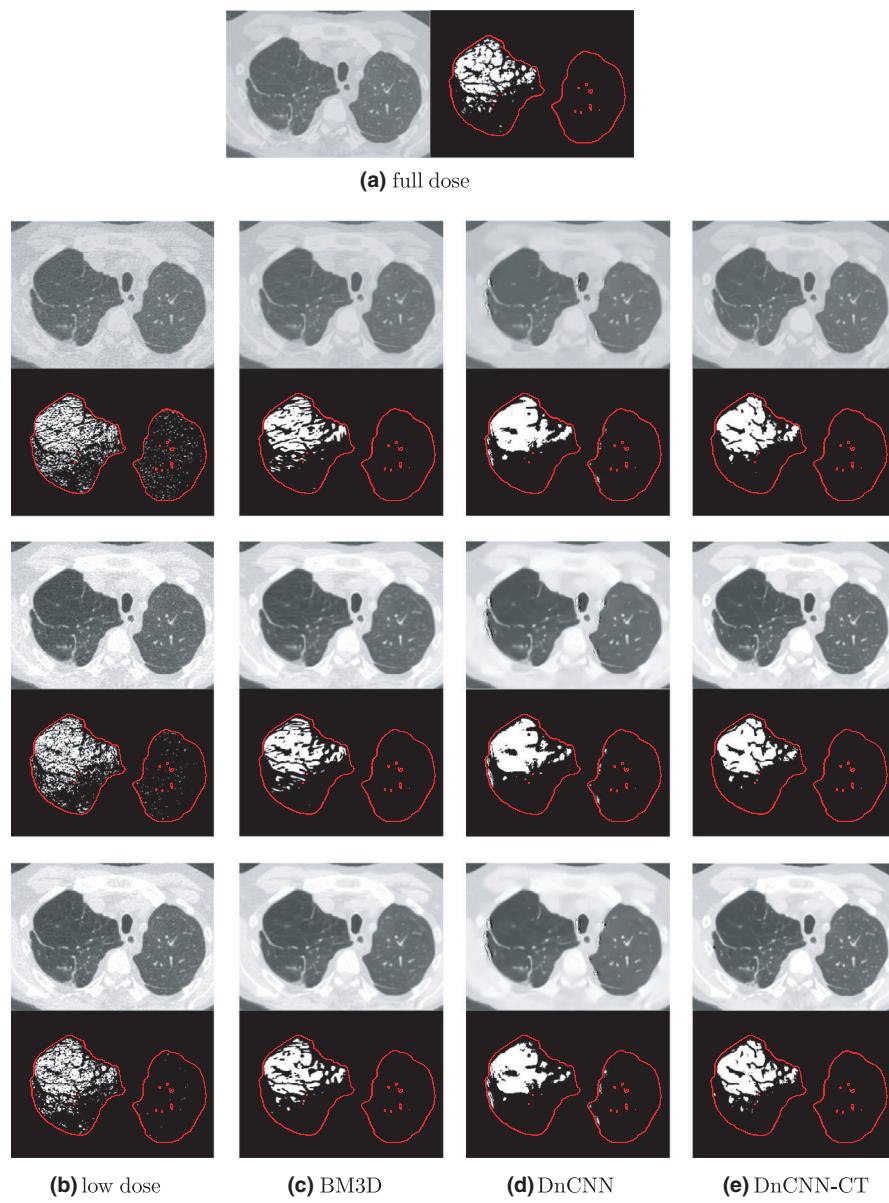


FIG. 6. A thoracic CT example with emphysema. Subfigure (a) illustrates the full-dose image slice with the corresponding emphysema mask. The other rows give the emphysema masks with dose reduction levels of 3%, 5% and 10% from top to bottom, and the low-dose image (b), denoised images using BM3D (c), the original DnCNN (d) and DnCNN-CT (e) from left to right. [Color figure can be viewed at wileyonlinelibrary.com]

3. EXPERIMENT SETUP AND RESULTS

3.A. Image denoising performance

3.A.1. PSNR and SSIM evaluation

In Figs. 3 and 4, we compare the denoising results in terms of PSNR and SSIM, respectively, among the BM3D denoising scheme, the original DnCNN and DnCNN-CT over different dose reduction levels. Both PSNR and SSIM are collected over 2D image slices as the denoising is performed in 2D. We can observe that both BM3D and DnCNN-CT significantly improve PSNR and SSIM while the original DnCNN yields less improvement or even high degradation. The image quality improvement is more prominent when the

dose reduction level is lower. For instance, tailoring DnNN with the CT data yields PSNR improvement of about 8 dB on average for our 3%-dose images, and SSIM increase by about 0.15.

3.A.2. Resolution characteristics

DnCNN, with spatially shared convolutional filters, has the beneficial property of locally invariant resolution change. However, global resolution is almost impossible to define or evaluate due to its high nonlinearity, which is similar to the nonlocal-mean-type filtering scheme, such as BM3D. As an effort to achieve some understanding, we qualitatively characterize the spatial resolution in Fig. 5, by examining the edge

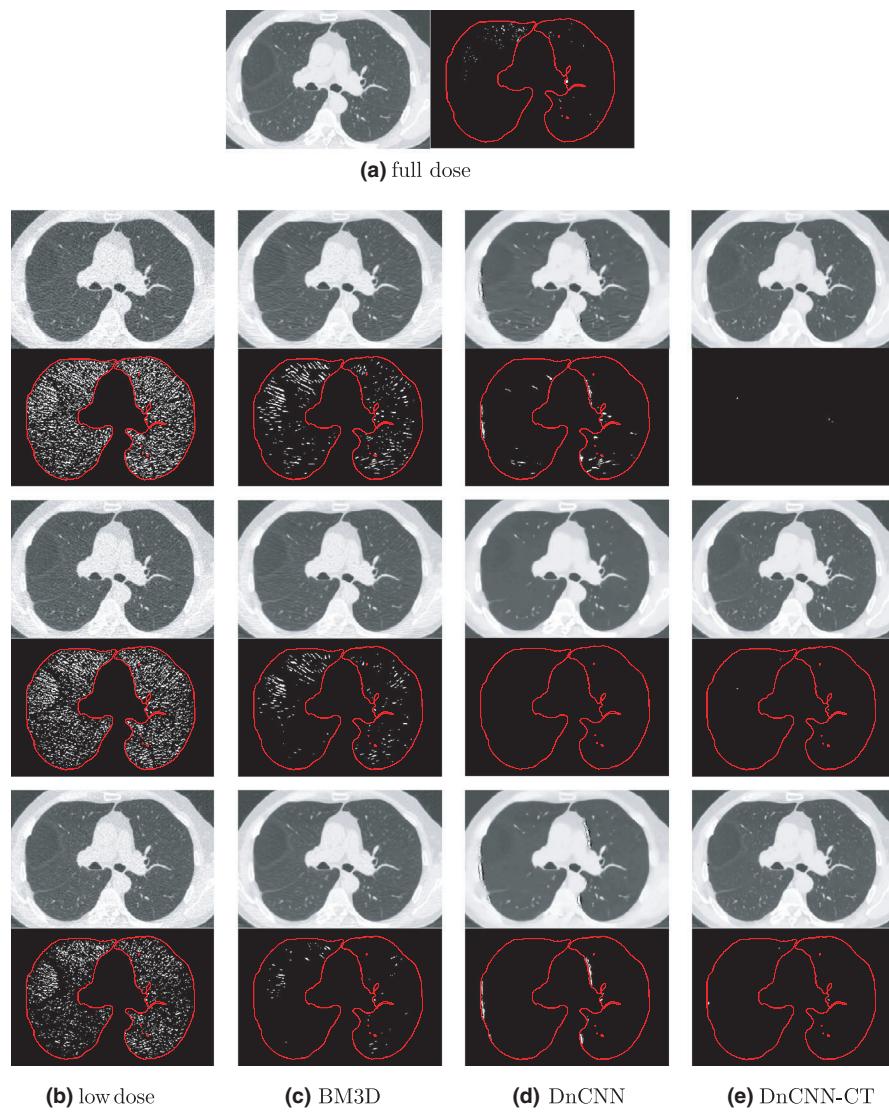


FIG. 7. A thoracic CT example without emphysema. Subfigure (a) illustrates the full-dose image slice with the corresponding emphysema mask. The other rows give the emphysema masks with dose reduction levels of 3%, 5%, and 10% from top to bottom, and the low-dose image (b), denoised images using BM3D (c), the original DnCNN (d), and DnCNN-CT (e) from left to right. [Color figure can be viewed at wileyonlinelibrary.com]

response in two different regions with edges at the mediastinum and lung/pleura boundaries, respectively. Images shown in this paper are displayed in lung window with window center of -600 HU and window width of 1500 HU to highlight the lung anatomy.

The illustrated region #1 across the mediastinum has low contrast (contrast level <120 HU) and the region #2 across the lung/pleura boundary shows much higher contrast (contrast level >800 HU). We can see from the edge profiles of both regions that the full-dose image gives distinct change at the two boundaries. The low-dose one shows high oscillation due to noise, which is successfully suppressed after denoising. For the low-contrast region #1, BM3D denoising results in an even flatter edge profile, increasing the 10–90% edge response distance by three voxels based on a coarse measure, which implies a reduced resolution. In contrast, the edge response distance collected after DnCNN-CT denoising is equal to that obtained from the full-dose image. For the high-

contrast region #2, the “sharp jump” (i.e., high resolution) across the boundary can be well recovered by any of the three denoising algorithms, yielding the same 10–90% edge response distance as the full-dose image.

3.B. Emphysema scoring

We applied the proposed denoising scheme to quantitative emphysema scoring in thoracic CT screening. Image slices with and without visible emphysema over different dose reduction levels are illustrated in Figs. 6 and 7, respectively. Their emphysema masks are generated based on the well-investigated measure RA950, indicating the percentage of lung voxels with attenuation coefficient lower than -950 HU. For ultra-low-dose images, the stochastic noise associated with dose reduction perturbs the attenuation value, similar to low-pass filtering of the probability distribution function (PDF), which artificially boosts the

TABLE I. Comparison of RA, PERC, mean lung intensity in HU, and KL divergence of the intensity distribution. For each testing case of 3% dose, the measures for full-dose reference are listed first, and then the deviation/KL divergence for each low-dose counterpart is shown. The (h,p) value is the test result and P -value resulting from paired t -test, compared between DnCNN-CT and other denoising methods.

		Percentile								
		RA970	RA950	RA930	RA910	PERC10	PERC15	PERC20	Mean	KL divergence
case 1	full dose	1.43e-3	1.47e-2	8.49e-2	2.46e-1	-927.53	-920.65	-914.86	-847.53	-
	3% dose	1.35e-1	1.82e-1	1.87e-1	1.12e-1	-57.18	-44.18	-34.24	-3.97	6.84e-1
	bm3d	1.56e-2	3.95e-2	5.21e-2	2.90e-2	-9.86	-7.09	-5.14	-3.08	4.54e-2
	DnCNN	8.48e-3	4.00e-3	-1.81e-2	-2.50e-2	3.35	3.09	2.75	-0.70	1.29e-2
	DnCNN-CT	-7.77e-4	-9.62e-3	-3.72e-2	-4.10e-2	6.34	5.33	4.44	-1.97	1.90e-2
case 2	full dose	9.09e-2	1.90e-1	2.82e-1	3.57e-1	-968.09	-958.03	-948.07	-812.46	-
	3% dose	1.08e-1	6.00e-2	2.20e-2	1.00e-3	-50.41	-33.75	-21.61	-4.69	2.84e-1
	bm3d	3.81e-2	1.50e-2	1.00e-3	-3.00e-3	-10.68	-6.30	-3.23	-3.89	7.53e-2
	DnCNN	1.21e-2	2.90e-2	2.90e-2	2.60e-2	-2.46	-4.15	-5.53	-2.94	1.28e-2
	DnCNN-CT	-1.56e-2	-1.00e-2	4.00e-3	1.70e-2	3.00	2.47	1.75	-4.91	7.37e-3
case 3	full dose	6.21e-4	1.74e-3	3.05e-3	4.85e-3	-858.78	-851.14	-844.69	-771.64	-
	3% dose	2.14e-2	3.60e-2	5.92e-2	9.35e-2	-50.40	-38.08	-28.59	-0.79	3.47e-1
	bm3d	2.59e-4	3.20e-4	1.83e-3	7.15e-3	-5.05	-2.12	-0.08	0.19	2.40e-2
	DnCNN	4.37e-3	4.56e-3	5.12e-3	6.65e-3	6.87	6.80	6.43	1.96	1.63e-2
	DnCNN-CT	-2.16e-4	-7.98e-4	-1.53e-3	-1.57e-3	-2.28	-2.84	-3.43	-7.09	9.69e-3
(h, p) t-test	3% dose	(1,0.00)	(1,0.00)	(1,0.00)	(1,0.00)	(0,1.00)	(0,1.00)	(0,1.00)	(1,0.00)	(1,0.00)
	bm3d	(1,0.00)	(1,0.00)	(1,0.00)	(0,9.34e-1)	(1,4.50e-4)	(1,0.00)	(1,0.00)	(0,1.00)	(0,9.77e-1)
	DnCNN	(1,0.00)	(1,0.00)	(1,0.00)	(1,0.00)	(1,2.24e-2)	(1,1.00e-4)	(1,1.18e-3)	(0,1.00)	(0,1.00)

We used one-sided pairwise t-test with null hypothesis being zero mean difference between the measure error of DnCNN-CT and that of the comparing method. The alternative hypothesis is that the measure error of DnCNN-CT is lower. Result $h = 1$ indicates the rejection of null hypothesis and $h = 0$ otherwise.

TABLE II. Comparison of RA, PERC, mean lung intensity in HU, and KL divergence of the intensity distribution. For each testing case of 5% dose, the measures for full-dose reference are listed first, and then the deviation/KL divergence for each low-dose counterpart is shown. The (h,p) value is the test result and P -value resulting from paired t -test, compared between DnCNN-CT and other denoising methods.

		Percentile								
		RA970	RA950	RA930	RA910	PERC10	PERC15	PERC20	Mean	KL divergence
case 1	full dose	1.43e-3	1.47e-2	8.49e-2	2.46e-1	-927.53	-920.65	-914.86	-847.53	-
	5% dose	8.64e-2	1.34e-1	1.47e-1	8.90e-2	-37.80	-28.97	-22.27	-2.60	3.98e-1
	bm3d	5.54e-3	1.77e-2	2.21e-2	6.00e-3	-3.85	-2.37	-1.36	-2.16	1.15e-2
	DnCNN	-1.43e-3	-1.46e-2	-7.50e-2	-1.02e-1	13.27	11.20	9.65	-0.84	7.22e-1
	DnCNN-CT	-9.66e-4	-1.16e-2	-4.76e-2	-6.10e-2	8.39	7.22	6.25	0.62	2.48e-2
case 2	full dose	9.09e-2	1.90e-1	2.82e-1	3.57e-1	-968.09	-958.03	-948.07	-812.46	-
	5% dose	8.21e-2	4.00e-2	9.00e-3	-4.00e-3	-32.35	-20.63	-12.27	-3.85	2.35e-1
	bm3d	2.11e-2	5.00e-3	-1.00e-3	-1.00e-3	-5.15	-2.57	-0.87	-3.06	2.89e-2
	DnCNN	-3.05e-2	3.00e-3	1.90e-2	2.50e-2	4.50	1.58	-0.86	-1.41	1.58e-2
	DnCNN-CT	-1.88e-2	-1.10e-2	0.00	1.10e-2	3.33	2.55	1.96	-2.42	7.94e-03
case 3	full dose	6.21e-4	1.74e-3	3.05e-3	4.85e-3	-858.78	-851.14	-844.69	-771.64	-
	5% dose	7.75e-3	1.51e-2	2.98e-2	5.61e-2	-33.58	-25.06	-18.56	-0.90	2.01e-1
	bm3d	-5.60e-5	-5.10e-4	-3.50e-4	1.61e-3	-0.24	1.28	2.25	0.05	8.52e-3
	DnCNN	-6.21e-4	-1.73e-3	-2.79e-3	-3.96e-3	15.08	13.12	11.52	0.81	
	DnCNN-CT	-4.57e-4	-1.10e-3	-1.77e-3	-2.87e-3	5.49	4.26	3.28	-0.82	1.12e-2
(h, p) t-test	5% dose	(1,0.00)	(1,0.00)	(1,0.00)	(1,0.00)	(0,1.00)	(0,1.00)	(0,1.00)	(1,0.00)	(1,0.00)
	bm3d	(1,0.00)	(0,5.23e-1)	(0,1.00)	(0,1.00)	(0,3.23e-1)	(1,0.00)	(1,3.19e-3)	(1,0.00)	(0,1.00)
	DnCNN	(1,0.00)	(0,7.00e-1)	(1,0.00)	(1,0.00)	(0,7.44e-1)	(0,4.52e-1)	(0,9.98e-1)	(1,0.00)	(1,0.00)

TABLE III. Comparison of RA, PERC, mean lung intensity in HU, and KL divergence of the intensity distribution. For each testing case of 10% dose, the measures for full-dose reference are listed first, and then the deviation/KL divergence for each low-dose counterpart is shown. The (h,p) value is the test result and p-value resulting from paired t-test, compared between DnCNN-CT and other denoising methods.

		RA970	RA950	RA930	RA910	Percentile			Mean	KL divergence
						PERC10	PERC15	PERC20		
case 1	full dose	1.43e-3	1.47e-2	8.49e-2	2.46e-1	-927.53	-920.65	-914.86	-847.53	-
	10% dose	3.88e-2	7.73e-2	9.51e-2	5.80e-2	-20.23	-15.30	-11.59	-1.53	1.48e-1
	bm3d	2.80e-4	3.00e-4	-8.60e-3	-2.00e-2	1.76	2.01	2.11	-1.11	2.27e-3
	DnCNN	8.47e-3	-2.70e-3	-5.67e-2	-1.04e-1	12.59	11.55	10.62	3.13	6.10e-2
	DnCNN-CT	-5.93e-4	-9.92e-3	-3.55e-2	-4.30e-2	5.99	5.18	4.54	0.71	1.29e-2
case 2	full dose	9.09e-2	1.90e-1	2.82e-1	3.57e-1	-968.09	-958.03	-948.07	-812.46	-
	10% dose	5.11e-2	2.00e-2	0.00	-7.00e-3	-16.33	-9.62	-4.86	-2.52	1.33e-1
	bm3d	-6.00e-4	-6.00e-3	-5.00e-3	-1.00e-3	0.35	1.01	1.31	-1.76	2.66e-3
	DnCNN	-5.90e-2	-2.90e-2	0.00	1.50e-2	9.70	6.50	3.85	0.79	4.14e-2
	DnCNN-CT	6.01e-3	-3.00e-3	6.00e-3	9.00e-3	4.25	1.92	0.34	-0.40	1.42e-2
case 3	full dose	6.21e-4	1.74e-3	3.05e-3	4.85e-3	-858.78	-851.14	-844.69	-771.64	-
	10% dose	1.72e-3	3.37e-3	8.55e-3	2.25e-2	-17.90	-13.00	-9.38	-0.63	7.87e-2
	bm3d	-2.18e-4	-8.03e-4	-1.30e-3	-1.47e-3	4.48	4.54	4.49	0.22	6.41e-3
	DnCNN	6.92e-3	6.86e-3	6.87e-3	6.75e-3	13.63	12.05	10.69	3.76	5.19e-2
	DnCNN-CT	-3.19e-4	-1.05e-3	-1.58e-3	-2.53e-3	4.72	3.97	3.42	0.36	7.29e-3
(h, p) t-test	10% dose	(1,0.00)	(1,0.00)	(1,0.00)	(1,0.00)	(0,1.00)	(0,1.00)	(0,1.00)	(1,0.00)	(1,0.00)
	bm3d	(0,1.00)	(0,1.00)	(0,1.00)	(0,1.00)	(0,8.43e-1)	(0,6.90e-2)	(0,4.11e-1)	(0,1.38e-2)	(0,1.00)
	DnCNN	(1,0.00)	(1,0.00)	(1,0.00)	(1,0.00)	(0,1.03e-1)	(0,4.37e-1)	(0,9.95e-1)	(1,0.00)	(1,0.00)

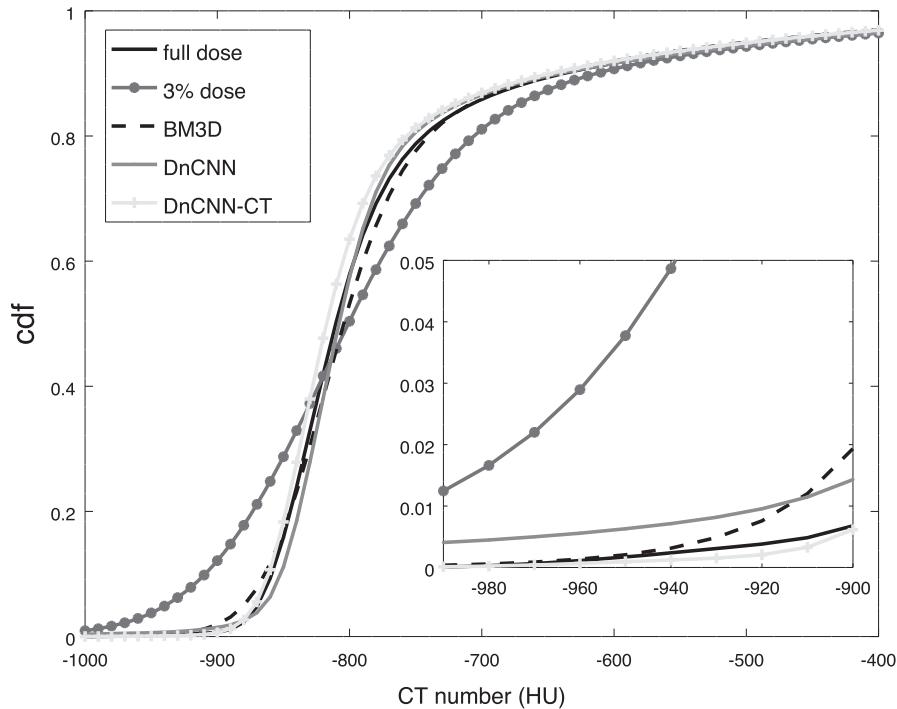


FIG. 8. The cdf of lung intensity for a full-dose image, the corresponding 3%-dose image, and those after denoising. KL divergence of the intensity distribution compared with the full-dose reference is 3.47e-1, 2.40e-2, 1.63e-2, and 9.69e-3, for 3%-low-dose images, the denoised images using BM3D, DnCNN, and DnCNN-CT, respectively.

tail portion of the pdf. We can observe that many more pixels are mistakenly marked as emphysema due to such “filtering” behavior, using emphysema masks obtained from full-dose images as the reference.

When visible emphysema exists as in Fig. 6, all denoising schemes effectively correct the false positives on voxels without emphysema while successfully detecting the abnormal area. The emphysema masks obtained from low-dose images

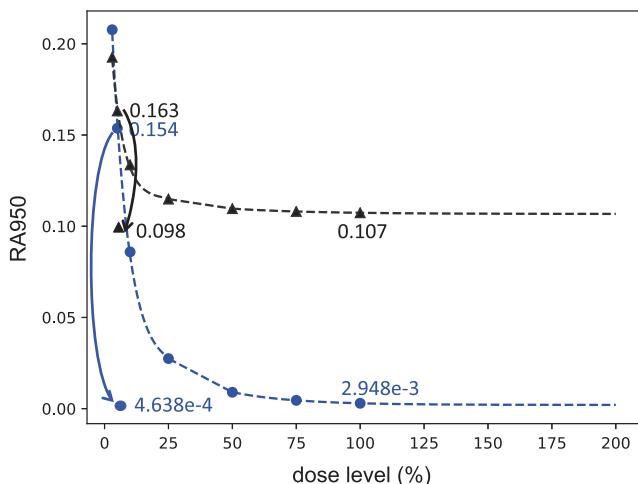


Fig. 9. RA950 value collected from images of 3%, 5%, 10%, 25%, 50%, 75%, and 100% dose, with exponential curve fitting results. Two cases with and without emphysema are illustrated. RA950 values corresponding to 100%-dose CT scan, 5%-dose counterpart and that after DnCNN-CT denoising are indicated for each curve. [Color figure can be viewed at [wileyonlinelibrary.com](#)]

after BM3D denoising agree best with that from full-dose one. In the case without emphysema as shown in Fig. 7, DnCNN-CT effectively reduces the high false-alarm probability while BM3D still has visible false detection, especially in the 3%-dose situation. In both cases, the original DnCNN yields artifacts after denoising, which results in labeling error.

The quantitative measures in emphysema scoring for the three testing cases are shown in Table I–III with dose reduction levels of 3%, 5%, and 10%, respectively. We define the error of a measure as the absolute value of difference between the measure obtained from (denoised) low-dose images and that from full-dose references. The results of paired t-test are also reported at 5% significance level, comparing the error after DnCNN-CT denoising against the input low-dose image slices or denoised slices by BM3D/DnCNN. We can see that DnCNN-CT works well on the ultra-low 3%-dose CT scans, with statistically significantly better RA and PERC measures than other approaches. For higher dose level of 10%, our DnCNN-CT demonstrates comparable performance with BM3D.

In addition, we illustrate the cumulative distribution function (cdf) of the CT intensity for an arbitrarily chosen image in Fig. 8. We can see that low-dose CT noise smears the intensity probability distribution and introduces more pixels artificially to the low-intensity tail. All cdf curves after denoising approach the full-dose reference. Magnified local view with intensity value below -900 HU demonstrates that DnCNN-CT yields good agreement with the full-dose reference except the smaller low-intensity tail due to over-smoothing. Though this will result in moderate mislabeling in the region when emphysema is scattered for abnormal cases, as illustrated in Fig. 6, it has little influence on the diagnostic decision. On the other hand, it manages to correct the severe overestimation in the non-emphysema cases.

3.C. Impact and clinical tolerance of over-smoothing associated with DnCNN-CT

It has been observed that there are over-smoothing behavior associated with DnCNN-CT results. However, we consider the value of DnCNN in denoising very low-dose CT scans and reaching a clinically sound decision, which outweighs its limitation of over-smoothing in image appearance. Figure 9 shows the RA950 measures over various dose reduction levels in the range of [3%, 100%]. Two CT cases are illustrated with and without emphysema, respectively. We can see for the case without emphysema, DnCNN-CT effectively reduces the discrepancy from more than 15% to less than 0.25% at 5%-dose level, compared to the full-dose reference. For the case with visible emphysema, the over-smoothing associated with DnCNN-CT introduces an underestimate of around 1%, compared to that from the full-dose reference, but does not affect clinical conclusions.

4. CONCLUSION AND DISCUSSION

This study achieved significant performance improvement by further tailoring an existing DnCNN natural image denoising model to the specific ultra-low-dose CT setup. We demonstrated its clinical value in the quantitative imaging network for emphysema scoring. The tailored DnCNN-CT demonstrated significant performance improvement in terms of image PSNR and SSIM while the original DnCNN yielded less gain or even image quality degradation. DnCNN-CT achieved comparable emphysema scoring performance with the state-of-the-art BM3D denoising. DnCNN-CT lightly underestimated the emphysema property due to over-smoothing, however, did not compromise its agreement to the diagnostic decision based on full-dose CT scans.

This work focused on “transferring” the current denoising model pre-trained on natural/synthetic images with Gaussian noise to the special context of CT reconstruction from low-dose acquisitions, and investigating its clinical impact. As a pilot study, the current investigation is limited in the number of samples available for both training and testing. We are in the process of collecting more projection data and running through the reconstruction pipeline to acquire further realistic pairing of low-dose and high-dose counterparts. Once a critical mass is accumulated, we will assess the feasibility of training a DnCNN from scratch²⁴ to understand the balance between the possible gain in effectiveness and the compromise due to model mismatch, resulting from transfer learning. A larger cohort will also enable more thorough statistical analysis, by performing various bootstrapping and cross-validation techniques.

ACKNOWLEDGMENT

We thank John Hoffman for his assistance with providing diagnostic data and supporting clinical applications.

^aAuthor to whom correspondence should be addressed. Electronic mail: druan@mednet.ucla.edu

REFERENCES

- National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med.* 2011;365:395–409.
- Brenner DJ. Radiation risks potentially associated with low-dose CT screening of adult smokers for lung cancer. *Radiology.* 2004;231:440–445.
- Brenner DJ, Hall EJ. Computed tomography—an increasing source of radiation exposure. *N Engl J Med.* 2007;357:2277–2284.
- Dabov K, Foi A, Katkovnik V, Egiazarian K. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Trans Image Process.* 2007;16:2080–2095.
- Lebrun M. An analysis and implementation of the BM3D image denoising method. *Image Process.* 2012;2:175–213.
- Elad M, Aharon M. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Trans Image Process.* 2006;15:3736–3745.
- Mairal J, Bach F, Ponce J, Sapiro G, Zisserman A. Non-local sparse models for image restoration. 2009 IEEE 12th International Conference on Computer Vision; 2009:2272–2279.
- Jain V, Seung S. Natural image denoising with convolutional networks. *Adv Neural Inf Process Syst.* 2009;769–776.
- Burger HC, Schuler CJ, Harmeling S. Image denoising: Can plain neural networks compete with BM3D?. Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, 2012;2392–2399.
- Plotz T, Roth S. Benchmarking Denoising Algorithms with Real Photographs. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017:1:6.
- He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016:770–778.
- Zhang K, Zuo W, Chen Y, Meng D, Zhang L. Beyond a gaussian denoiser: residual learning of deep CNN for image denoising. *IEEE Trans Image Process.* 2017;26:3142–3155.
- Wang J, Wang S, Li L, Fan Y, Lu H, Liang, Z. Virtual colonoscopy screening with ultra low-dose CT and less-stressful bowel preparation: a computer simulation study. *IEEE Trans Nucl Sci.* 2008;55:2566–2575.
- Hoffman J, Young S, Noo F, McNitt-Gray, M. Technical Note: FreeCT_wFBP: a robust, efficient, open-source implementation of weighted filtered backprojection for helical, fan-beam CT. *Med Phys.* 2016;43:1411–1420.
- Mayo JR, Kim KI, MacDonald SL, et al. Reduced radiation dose helical chest CT: effect on reader evaluation of structures and lung findings. *Radiology.* 2004;232:749–756.
- Yuan R, Mayo JR, Hogg JC, et al. The effects of radiation dose and CT manufacturer on measurements of lung densitometry. *CHEST J.* 2007;132:617–623.
- Parr DG, Stolk BC, Stolk J, Stockley RA. Validation of computed tomographic lung densitometry for monitoring emphysema in α 1-antitrypsin deficiency. *Thorax.* 2006;61:485–490.
- Hoessein FAM, de Hoop B, Zanen P, et al. CT-quantified emphysema in male heavy smokers: association with lung function decline. *Thorax.* 2011;66:782–787.
- Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. International Conference on Machine Learning, 2015.
- Zabić S, Wang Q, Morton T, Brown KM. A low dose simulation tool for CT systems with energy integrating detectors. *Med Phys.* 2013;40:031102.
- Young S, Kim HG, Ko MM, Ko WW, Flores C, McNitt-Gray M. Variability in CT lung-nodule volumetry: effects of dose reduction and reconstruction methods. *Med Phys.* 2015;42:2679–2689.
- Choo JY, Goo JM, Lee CH, Park CM, Park SJ, Shim MS. Quantitative analysis of emphysema and airway measurements according to iterative reconstruction algorithms: comparison of filtered back projection, adaptive statistical iterative reconstruction and model-based iterative reconstruction. *Eur Radiol.* 2014;24:799–806.
- Nishio M, Koyama H, Ohno Y, et al. Emphysema quantification using ultralow-dose CT with iterative reconstruction and filtered back projection. *Am J Roentgenol.* 2016;206:1184–1192.
- He K, Girshick R, Dollár P. Rethinking ImageNet Pre-training. *arXiv preprint arXiv:1811.08883*, 2018.