

Multislice input for 2D and 3D residual convolutional neural network noise reduction in CT

Zhongxing Zhou^{ORCID}, Nathan R. Huber, Akitoshi Inoue^{ORCID},
Cynthia H. McCollough^{ORCID}, and Lifeng Yu^{ORCID}*

Mayo Clinic, Department of Radiology, Rochester, Minnesota, United States

Abstract

Purpose: Deep convolutional neural network (CNN)-based methods are increasingly used for reducing image noise in computed tomography (CT). Current attempts at CNN denoising are based on 2D or 3D CNN models with a single- or multiple-slice input. Our work aims to investigate if the multiple-slice input improves the denoising performance compared with the single-slice input and if a 3D network architecture is better than a 2D version at utilizing the multislice input.

Approach: Two categories of network architectures can be used for the multislice input. First, multislice images can be stacked channel-wise as the multichannel input to a 2D CNN model. Second, multislice images can be employed as the 3D volumetric input to a 3D CNN model, in which the 3D convolution layers are adopted. We make performance comparisons among 2D CNN models with one, three, and seven input slices and two versions of 3D CNN models with seven input slices and one or three output slices. Evaluation was performed on liver CT images using three quantitative metrics with full-dose images as reference. Visual assessment was made by an experienced radiologist.

Results: When the input channels of the 2D CNN model increases from one to three to seven, a trend of improved performance was observed. Comparing the three models with the seven-slice input, the 3D CNN model with a one-slice output outperforms the other models in terms of noise texture and homogeneity in liver parenchyma as well as subjective visualization of vessels.

Conclusions: We conclude that multislice input is an effective strategy for improving performance for 2D deep CNN denoising models. The pure 3D CNN model tends to have a better performance than the other models in terms of continuity across axial slices, but the difference was not significant compared with the 2D CNN model with the same number of slices as the input.

© 2023 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: [10.1117/1.JMI.10.1.014003](https://doi.org/10.1117/1.JMI.10.1.014003)]

Keywords: deep convolutional neural network; noise reduction; multislice input.

Paper 22142GRR received Jun. 7, 2022; accepted for publication Jan. 9, 2023; published online Jan. 31, 2023.

1 Introduction

It is best practice in clinical computed tomography (CT) to use as low as reasonably achievable (ALARA) radiation dose in each exam without compromising diagnostic image quality.¹ However, the reduction of radiation dose often comes at the cost of increasing noise and artifacts in the reconstructed images, which may negatively affect the diagnostic performance. To address this issue, multiple noise reduction techniques have been developed: these include projection data filtering,^{2,3} iterative reconstruction (IR) algorithms,⁴⁻⁷ and denoising on CT images after reconstruction.⁸⁻¹⁰ Among these methods, various IR methods have been widely deployed on clinical scanners for radiation dose reduction or image quality improvement. It was shown that, at most, a 20% to 30% dose reduction is allowed without sacrificing diagnostic performance

*Address all correspondence to Lifeng Yu, Yu.Lifeng@mayo.edu

for low-contrast lesion detection tasks.¹¹ One common problem associated with these IR methods is the perceived change of noise texture and relatively long computation time.¹²

In recent years, deep learning-based methods using convolutional neural networks (CNN) have received increasing attention mainly due to their potential of preserving natural texture after substantial noise reduction and high computational efficiency. Various deep CNN methods have been proposed to reduce the noise of CT images after image reconstruction. Chen et al. introduced a lightweight CNN-based framework with a single-slice input for low-dose CT images denoising.¹³ To improve the CNN denoising performance, Yang et al. adopted Wasserstein distance and perceptual loss for a generative adversarial network.¹⁴ For the same purpose of noise suppression and structural preservation, Chen et al.¹⁵ developed a residual encoder–decoder CNN network for CT image denoising. To capture a wide range of spatial information both within CT slices and between CT slices, 2D CNN models with a three-slice input were adopted in low-dose CT imaging, which was proposed for natural image classification with three input channels.¹⁶ This strategy was further investigated in 3D CNN-based low-dose CT imaging. One version of 3D CNN models utilized the network architecture with fully 3D CT image as the input and output^{17–19} and achieved better performance in both edge-preservation and noise-artifact suppression than the 2D CNN model with a single-slice input. Another version of 3D CNN models is a hybrid 2D/3D network with a 2D output,²⁰ which was shown to be able to save considerable computational time and achieve a better performance in suppressing image noise and preserving subtle structures than the 2D CNN model with a single-slice input. Presumably a multislice input in either 2D or 3D CNNs has the potential to incorporate spatial information from adjacent slices.^{16–20} In previous works, CT image denoising performance has been compared between 2D or 3D CNN models.^{17–20} But in these works, only the 2D CNN models with single-slice inputs were adopted for comparison. The benefits of including three or more CT slices for model training, especially in 2D CNN models, have not been clearly demonstrated. In addition, there remains a question of whether a 3D CNN model is better than its 2D counterpart with the same number of slices as the input.

The purpose of this work is to answer these two questions by evaluating the performance among 2D CNN models with different numbers of input slices and among 2D and 3D models with the same number of input slices. We include two versions of 3D CNN models in this study; the first has a 3D volume input and 2D output,²⁰ and the second one is a pure 3D model with an output that is also a 3D volume.¹⁸

2 Methods

2.1 2D/3D Residual CNN Framework

We employed a recently developed residual CNN denoiser^{16,21} as the basic network architecture for the 2D CNN denoiser. In the previous work, we used three CT slices as the three-channel input of the 2D residual CNN model. Here we varied the number of input channels to determine whether it affects the denoising performance. The 2D residual CNN architecture is shown in Fig. 1. The CNN inputs were first standardized (subtracted by the mean value and divided by the standard deviation) and then subjected to initial layers that generated 128 feature maps using 2D convolutional layers. The feature maps were further processed by a series of 2D residual blocks, each of which consisted of repeated layers of 2D convolutional, batch normalization, and rectified linear unit activation. The filter number of each 2D convolutional layer was set to 128, and the kernel size of all layers was set to 3×3 with a stride of one in each dimension. Then the output of residual blocks was projected back to a single-channel image using a single convolutional layer with one filter. This single-channel image was the estimated noise, which was further subtracted from the central input slice to get the final denoising result. Mirror padding in the axial dimensions was employed in all of the convolutional layers for all CNN models in this study. Hereafter, residual 2D CNN noise reduction with an input of one slice, three slices, and seven slices will be referred to as ResNet2D_1Slice, ResNet2D_3Slices, and ResNet2D_7Slices, respectively.

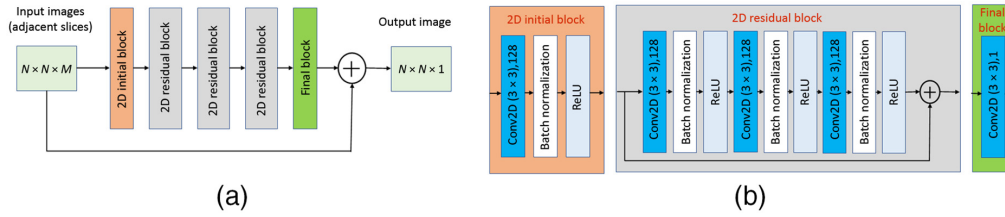


Fig. 1 Architecture of a residual 2D CNN denoiser. (a) Global structure of the network containing a 2D initial block, three 2D residual blocks, and a final block. (b) Details of the convolutional layers and transformations used within each block. Conv2D, two-dimensional convolutional layer; N , arbitrary image size; M , number of slices (1, 3, and 7), ReLU, rectified linear units.

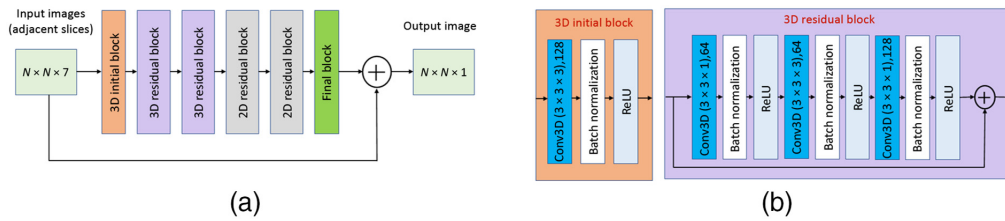


Fig. 2 Architecture of the first version of the residual 3D CNN denoiser. (a) Global structure of the network containing a 3D initial block, two 3D residual blocks, two 2D residual blocks, and a final block. (b) Details regarding the convolutional layers and transformations used within each block. The structures of 2D residual blocks and the final block are the same as those in Fig. 1.

To study the impact of 3D residual CNN, two versions of the architecture were designed. The architecture of the first version (ResNet3D-v1) is shown in Fig. 2. We replaced the 2D initial block and the first 2D residual block with a 3D initial block and two 3D residual blocks, respectively. To reduce the computational burden and alleviate the overfitting problem, we employed a bottleneck structure to reduce the dimension of feature maps in the 3D residual blocks [see Fig. 2(b)]. No padding was applied to the slice dimension before each 3D convolution. Hence after the second 3D residual block, the input 3D rectangular volume of $N \times N \times 7$ voxels reduced to $N \times N \times 1$ voxels, which can be squeezed to the 2D input of the next 2D residual block. Note that to have a fair comparison with the ResNet2D_7Slices model, which has the best denoising performance among the three ResNet2D models, we only consider the scenario of the ResNet3D models with the seven-slice input. Better denoising performance of the 3D CNN model can be achieved using seven slices than fewer slices as the input according to the results in Ref. 17.

The second version of the 3D residual CNN (ResNet3D-v2) architecture is shown in Fig. 3. We replaced the remaining 2D convolution layers in ResNet3D-v1 with 3D ones to make the model a pure 3D one. Mirror padding on the slice dimension was employed before the 3D convolutions with a convolution kernel of $3 \times 3 \times 3$ for the first two 3D residual blocks but not for

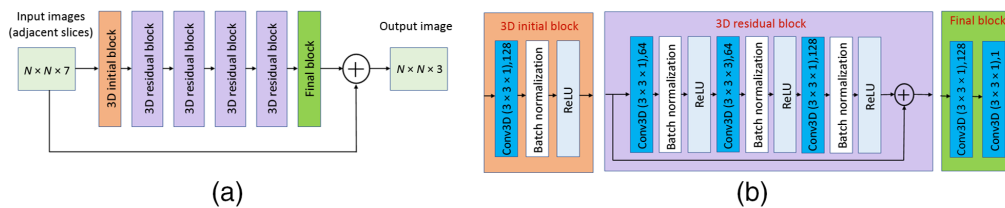


Fig. 3 Architecture of the second version of the residual 3D CNN denoiser. (a) Global structure of the network containing a 3D initial block, four 3D residual blocks, and a final block. (b) Details regarding the convolutional layers and transformations used within each block.

the next two ones. Hence, the size of the output is equal to the original input size ($N \times N \times 7$) after the second 3D residual block and is reduced to $N \times N \times 3$ voxels after the fourth residual block. A kernel size of 1 in the slice dimension was adopted for the 3D convolutions in the final block, so the final output of ResNet3D-v2 is a 3D volume of $N \times N \times 3$ voxels. Only the center slice of each output 3D volume is retained for each anatomy position. The final 3D volume of slices is formed from all CNN denoised outputs.

The residual step with a global shortcut connection from the input to output is on top of those with local shortcut connections within each 2D or 3D residual block. Relatively dense local shortcut connections facilitate deep residual learning.²⁰ The global shortcut connection results in the network predicting the noise residual and then removing this noise from the input directly,¹⁵ which can preserve structural details as well as make the training process stably and quickly.²¹

2.2 Training/Validation and Testing Data Set

CNN noise reduction models were trained using images from the Mayo/American Association of Physicists in Medicine (AAPM) Low-Dose CT Grand Challenge dataset.²² This dataset consists of full-dose (FD) abdominal CT images that were taken from 30 patients and the corresponding simulated quarter-dose (QD) CT images. All training data were reconstructed with a field of view of 340 mm, a medium smooth kernel (B30), and slice thickness and interval of 1.0 and 0.8 mm, respectively. We applied the overlapped sliding window with a sliding size of $32 \times 32 \times 1$ to obtain 2D image patches of size 64×64 and with a sliding size of $32 \times 32 \times 3$ to obtain 3D image patches of size $64 \times 64 \times 3$ and $64 \times 64 \times 7$. For each CNN denoiser, 512,000 matched QD and FD patches were randomly selected from 17 patients of this dataset for training, and 64,000 matched QD and FD patches were randomly selected from another five patients to validate the performance of the trained models. During the training phase, the adjacent QD patches were used as inputs of 3D residual models or 2D ones with multislice input, and the corresponding central single or three slices from the FD patches was used as the CNN target. The Adam optimizer²³ with a descending learning rate from 0.001 to 0.00001 was used to train the residual CNN models with a minibatch of 16 image patches for each iteration. We employed pixel-wise mean-square-error between CNN output and FD images as the loss function during optimization. The CNN training was performed on a NVIDIA Tesla M40 GPU with 12 GB memory.

Eight patient exams from the grand challenge dataset were reserved for testing the performance of residual CNN models. During testing phase, the five residual CNN models were applied to the QD images without or with adjacent slices along the scan direction (512×512 , $512 \times 512 \times 3$, and $512 \times 512 \times 7$).

2.3 Model Evaluation

The denoising performance of three 2D models with the one-, three-, and seven-slice input (ResNet2D_1Slice, ResNet2D_3Slices, ResNet2D_7Slices) and two 3D models with the seven-slice input (ResNet3D-v1, ResNet3D-v2) were evaluated using quantitative metrics and a radiologist's visual assessment in comparison with QD images reconstructed with filtered-backprojection (FBP) and IR methods and FD images reconstructed with FBP method, named QD + FBP, QD + IR, and FD + FBP images, respectively. FBP reconstructed axial images with 512×512 matrix size at QD from eight reserved patients unseen during training (around 500 axial images per patient) were denoised by five trained CNN models, respectively. Three quantitative metrics, including root-mean-square error (RMSE), peak-signal-to-noise ratio (PSNR), and structural similarity index measure (SSIM) were calculated between each FD + FDP axial image and the corresponding CNN denoised axial image at the same anatomical position from each of the five CNN models.

The RMSE is a measure of the residuals between the model predictions and the target image. PSNR is the ratio of signal power to noise power in the image (in decibel unit), and SSIM measures perceptual similarity based on luminance, contrast, and structure. Paired *t*-tests were performed to compare the quantitative metrics of different CNN models ($p < 0.05$ was considered

significantly different) across all processed axial images of eight testing patients. Because RMSE and PSNR are not directly related to the perceptual quality of the denoised images, images from each model were also visually assessed by an experienced radiologist (11 years' experience) by ranking all five CNN models, along with the other three conditions: QD + FBP, QD + IR, and FD + FBP, from best to worst for visualizing small anatomical structures, such as branch of mesenteric vein, left gastric vein, intrahepatic portal vein, posterior superior pancreaticoduodenal vein, anterior superior pancreaticoduodenal artery, and jejunal artery.

Images processed by each of the eight algorithms were displayed simultaneously on a multi-viewer in a single sitting in a darkened reading room. The radiologist blinded to the conditions of the study reviewed the axial and coronal images of all eight cases. The radiologist was asked to rank images from each algorithm in terms of image noise texture (graininess/homogeneity of the image in the liver parenchyma), sharpness (subjective impression on the sharpness of the edge of the hepatic vessels), and artifacts (artificial zig-zag linear structure or streaking artifacts originated from iodine or bony structures).²⁴ For structure preservation, the radiologist was instructed to determine if the small vessels in the liver can be clearly identified as expected in normal sectional anatomy. For continuity between slices, the radiologist was asked to evaluate the smoothness of the intrahepatic vessels while scrolling images back and forth. A five-point Likert scale [(1) poor, (2) adequate, (3) good, (4) very good, (5) excellent] was rated subjectively by comparing with clinical routine abdominal CT images.

3 Results

The parameter counts and training time per epoch and the inference time of one 512×512 image for each CNN model are shown in Table 1. Including more slices as the input does not significantly impact the training or inference time of the 2D residual CNN models. Although the five models have a similar number of parameters, the 3D models consumed significantly more time for training and inference, especially for the pure 3D model (ResNet3D-v2). After approximately 35 epochs, the training was stopped due to no improvement in the validation loss for any of the CNN models.

Figure 4 shows the quantitative evaluation results among 2D and 3D residual CNNs. The means \pm SDs (average scores \pm standard deviations) of PSNR, SSIM, and RMSE across all slices in the processed eight patient CT volumes (around 500 slices per volume) are plotted. The ResNet2D_3Slices model had higher PSNR and SSIM and lower MSE than the ResNet2D_1Slice model ($p < 0.001$ for all comparisons). The ResNet2D_7Slices model performed even better than the ResNet2D_3Slices model as confirmed by these three metrics ($p < 0.001$ for PSNR and MSE, $p = 0.004$ for SSIM), demonstrating the benefit of the multislice input. The ResNet3D-v1 model achieved significantly better performance than the ResNet2D_3Slices model ($p < 0.001$ for PSNR and MSE, $p = 0.003$ for SSIM). The ResNet3D-v2 model had significantly higher PSNR and lower MSE ($p < 0.001$) but no significant difference in SSIM ($p = 0.12$) compared with the ResNet2D_3Slices model. Among the three CNN models with the seven-slice input, no significant difference was found in terms of PSNR, SSIM, and RMSE ($p > 0.1$ for all comparisons).

Table 1 Training time and parameter count for the five CNN models.

	ResNet2D_1Slice	ResNet2D_3Slices	ResNet2D_7Slices	ResNet3D-v1	ResNet3D-v2
Parameter counts (million)	1.336	1.338	1.343	1.412	1.188
Training time per epoch (h)	0.40	0.43	0.45	1.2	4.5
Inference time per image (s)	0.16	0.19	0.20	0.54	1.25

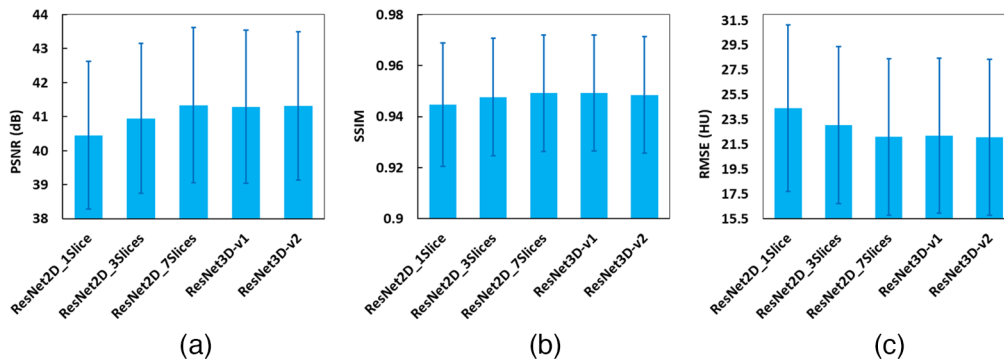


Fig. 4 Quantitative comparison among 2D and 3D residual CNNs in terms of (a) PSNR, (b) SSIM, and (c) RMSE.

Three image examples are shown in Figs. 5–7. Figure 5 compares the QD images reconstructed by FBP and IR, full dose (FD) images reconstructed by FBP, and the five versions of CNN denoising of the QD + FBP images (ResNet2D_1Slice, ResNet2D_3Slices, ResNet2D_7Slices, ResNet3D-v1, and ResNet3D-v2) from a patient case in the Mayo/AAPM Low-dose CT Grand Challenge data library (case no. L123). Images at a slice location containing many small anatomical structures were displayed: a branch of mesenteric vein (red arrow), left gastric vein (blue arrow), anterior superior pancreaticoduodenal artery (yellow arrow), posterior superior pancreaticoduodenal vein (green arrow), and jejunal artery (green circle). The performance of five CNN models was assessed visually by an experienced radiologist. Overall, the radiologist’s ranking order (from best to worst) for visualizing those small structures was (1) ResNet3D-v1, (2) ResNet2D_7Slices and ResNet3D-v2 (comparable),

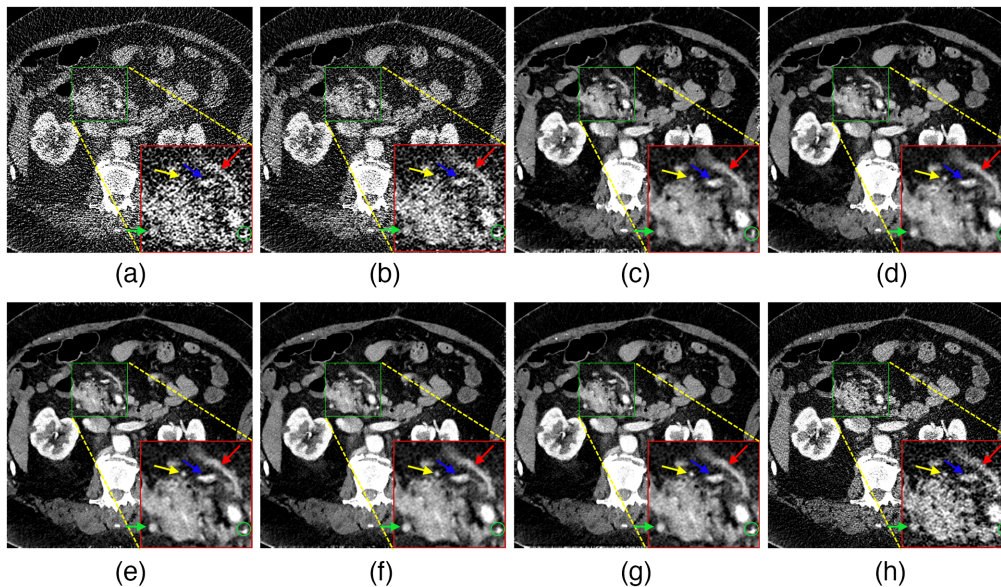


Fig. 5 Example image comparing the FBP and IR QD, FBP FD, and the five versions of CNN denoising of the QD images (ResNet2D_1Slice, ResNet2D_3Slices, ResNet2D_7Slices, ResNet3D-v1, and ResNet3D-v2) from a patient case in the Mayo/AAPM Low-dose CT Grand Challenge data library (case no. L123). (a) QD + FBP, (b) QD + IR, (c) ResNet2D_1Slice, (d) ResNet2D_3Slices, (e) ResNet2D_7Slices, (f) ResNet3D-v1, (g) ResNet3D-v2, and (h) FD + FBP. The red rectangles indicate zoomed-in ROIs of the green rectangles on the original image, with some small vessels labeled as follows: branch of mesenteric vein (red arrow), left gastric vein (blue arrow), anterior superior pancreaticoduodenal artery (yellow arrow), posterior superior pancreaticoduodenal vein (green arrow), and jejunal artery (green circle). The display window of this slice is [40, 300] HU.

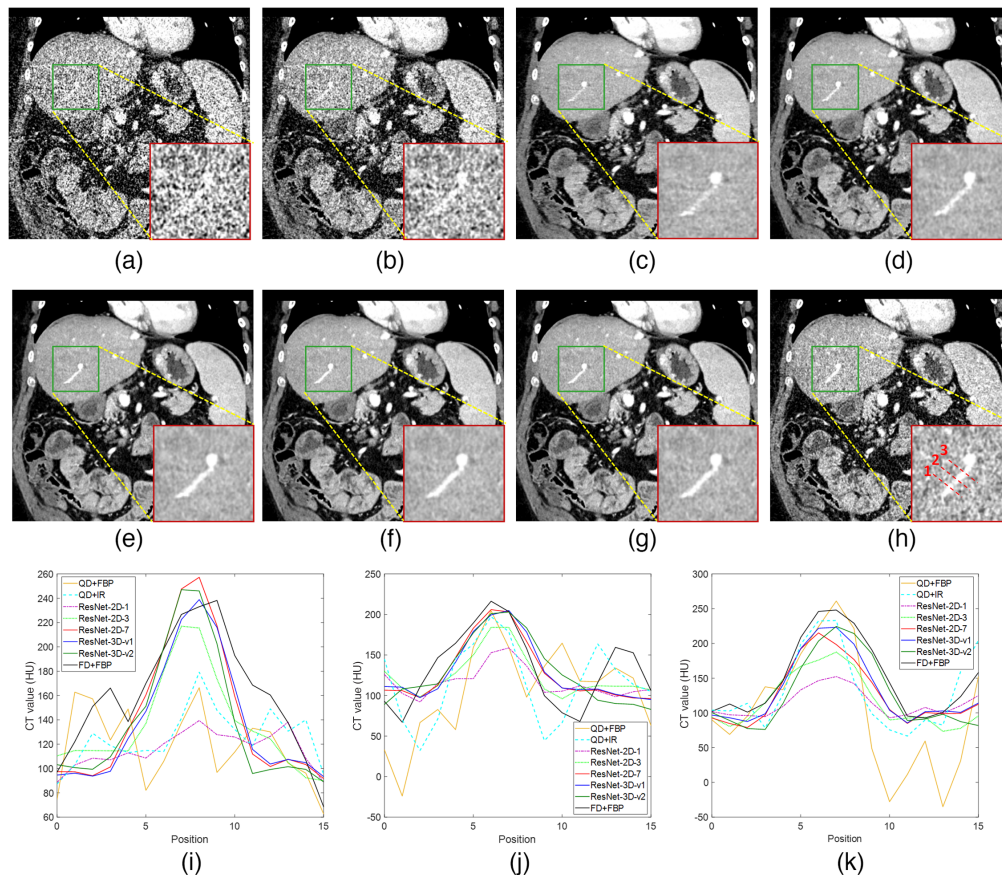


Fig. 6 Coronal reformatted image comparing the FBP and IR QD, FBP FD, and the five versions of CNN denoising of the QD images (ResNet2D_1Slice, ResNet2D_3Slices, ResNet2D_7Slices, ResNet3D-v1, and ResNet3D-v2) from a patient case in the Mayo/AAPM Low-dose CT Grand Challenge data library (case nno. L061). (a) QD + FBP, (b) QD + IR, (c) ResNet2D_1Slice, (d) ResNet2D_3Slices, (e) ResNet2D_7Slices, (f) ResNet3D-v1, (g) ResNet3D-v2, and (h) FD + FBP. The red rectangles indicate zoomed-in ROIs of the green rectangles in the original images. Three profiles were plotted in (i)–(k) along a direction vertical to the intrahepatic portal vein [labeled in dashed lines in (h)]. The display window of this slice is [40, 300] HU.

(3) ResNet2D_3Slices, and (4) ResNet2D_1Slice. The ResNet2D_1Slice model showed the posterior superior pancreaticoduodenal vein with lower enhancement than pancreatic parenchyma despite all five small vessels being visualized. The ResNet2D_7Slices model better depicted the branch of mesenteric vein and posterior superior pancreaticoduodenal vein with more smooth edges than the ResNet2D_3Slices model. The two ResNet3D models have a similar performance to the ResNet2D_7Slices model in terms of the visibility of vessels. Focusing on the pancreatic parenchyma, the ResNet3D-v1 model had less noise than the ResNet2D_7Slices and ResNet3D-v2 models. The original QD + FBP and QD + IR images were too noisy to visualize those small vessels for a diagnostic purpose. All CNN denoising results were visually better than the original FD + FBP image.

Figure 6 shows another example in a coronal reformat (case no. L061). As shown in the zoomed-in images, the intrahepatic portal vein is well traced with the smoothest edge in the ResNet3D-v1 [Fig. 6(f)], whereas the two denoising images from ResNet2D_1Slice and ResNet2D_3Slices models show the vein with obvious mosaic-like and discontinuous appearance. The overall radiologist’s ranking order (from best to worst) was (1) ResNet3D-v1, (2) ResNet3D-v2, (3) ResNet2D 7Slices, (4) ResNet2D_3Slices, and (5) ResNet2D_1Slice. Three profiles vertical to the intrahepatic portal vein [labeled with three dashed lines in Fig. 6(h)] were plotted [Figs. 6(i)–6(k)]. It can be seen that ResNet3D-v1, ResNet3D-v2, and ResNet2D with the seven-slice input better preserved the profiles from the corresponding FD

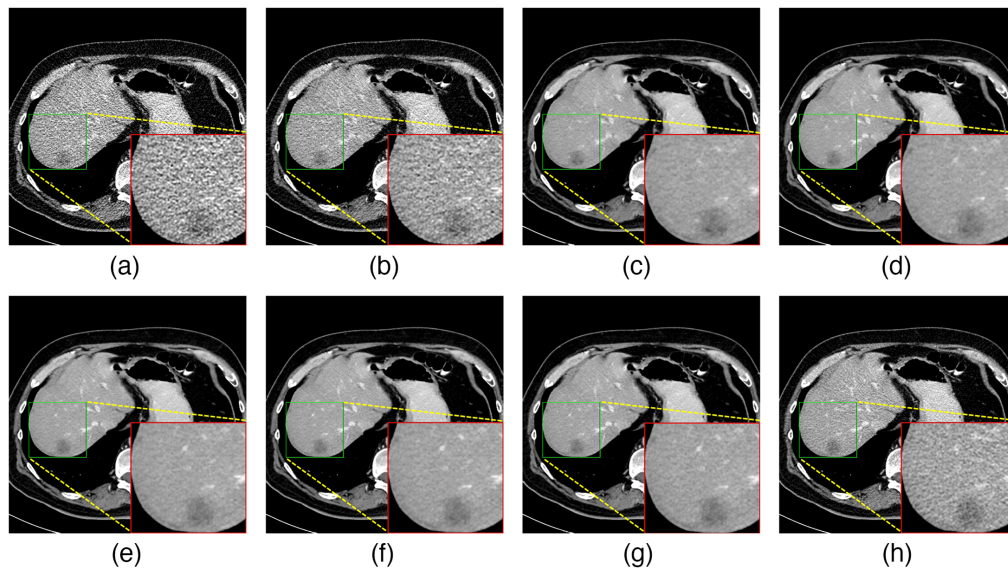


Fig. 7 Example image comparing the FBP and IR QD, FBP FD, and the five versions of CNN denoising of the QD images (ResNet2D_1Slice, ResNet2D_3Slices, ResNet2D_7Slices, ResNet3D-v1, and ResNet3D-v2) from a patient case in the Mayo/AAPM Low-dose CT Grand Challenge data library (case no. L136). (a) QD + FBP, (b) QD + IR, (c) ResNet2D_1Slice, (d) ResNet2D_3Slices, (e) ResNet2D_7Slices, (f) ResNet3D-v1, (g) ResNet3D-v2, and (h) FD + FBP. The red rectangles indicate zoomed-in ROIs of the green rectangles on the original image, where a hepatic metastasis is presented. The display window of this slice is [40, 300] HU.

image than the other models. The ResNet3D-v1 and ResNet3D-v2 slightly outperformed the ResNet2D-7Slices in terms of noise texture and homogeneity in liver parenchyma as well as subjective visualization of the target vein.

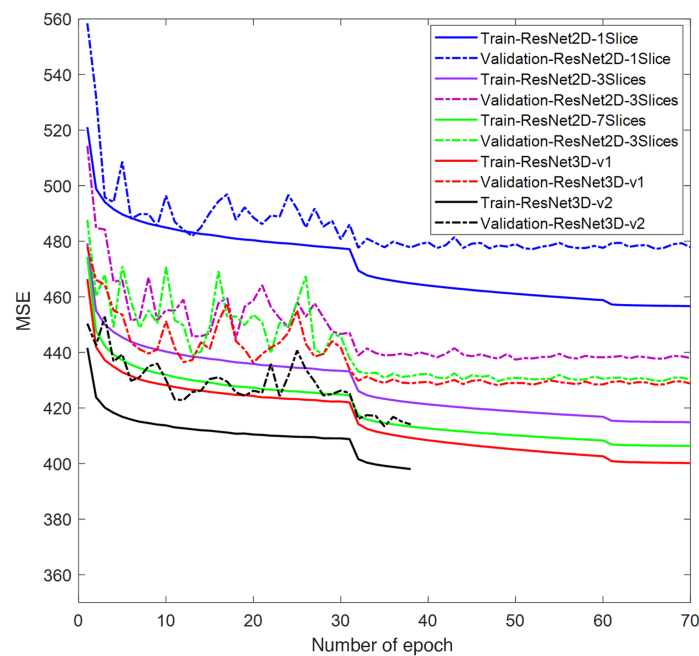
In Fig. 7, the black dot artifact in the hepatic metastasis is obviously seen in the ResNet2D_1Slice and ResNet2D_3Slices models, despite there being a similar overall image quality among the five denoising images based on the subjective evaluation of the radiologist. A similar trend of noise reduction was noted in liver images as pancreas images from all CNN denoising algorithms, ResNet3D-v1 is less noisy than ResNet2D_7Slices and ResNet3D-v2. The overall radiologist's ranking order (from best to worst) was (1) ResNet3D-v1, (2) ResNet2D_7Slices and ResNet3D-v2 (comparable), (3) ResNet2D_3Slices, and (4) ResNet2D_1Slice.

In the large patients, artifacts described as nodular or patchy low- or high attenuating areas in the liver parenchyma, which mimic liver lesions, were marked despite less noise texture in ResNet2D_1Slice and ResNet2D_3Slices. The artifact was linked to a lower score of overall image quality in the radiologist's interpretation. For each method, the scores were reported as means \pm standard deviations (SDs) across all eight patient cases. The FD + FBP images were used as the reference standard. The student *t*-test with $p < 0.05$ was performed, and the statistical results are provided in Table 2.

For all six image quality criteria, the three CNN models with the seven-slice input (ResNet2D_7Slices, ResNet3D-v1, and ResNet3D-v2) had higher scores than QD + FBP, QD + IR, and ResNet2D_1Slice ($p < 0.05$ for all comparisons), but no significant difference was found among these CNN models with the seven-slice input ($p > 0.05$ for all comparisons). The ResNet2D_1Slice and ResNet2D_3Slices models had scores superior to QD + FBP and QD + IR in terms of noise texture, structure preservation, sharpness, and overall image quality ($p < 0.05$ for all comparisons), but no significant difference or even a worse performance was found in terms of artifacts and continuity between slices. For some cases, when the intra- and extrahepatic vessels were focused, ResNet3D-v2 had a better performance in continuity between slices than ResNet2D_7Slices and ResNet3D-v1, but there was no statistically significant difference for all cases. In terms of artifacts, the two 3D CNN models performed similarly in comparison with the ResNet2D_7Slices model ($p > 0.05$). ResNet3D-v2 was ranked the best

Table 2 Subjective image quality scores for different algorithms (mean \pm SDs).

	QD + FBP	QD + IR	ResNet2D_1Slice	ResNet2D_3Slices	ResNet2D_7Slices	ResNet3D-v1	ResNet3D-v2	FD+FBP
Noise texture	1.25 \pm 0.46	1.88 \pm 0.35	3.63 \pm 0.52	3.88 \pm 0.35	4.25 \pm 0.46	4.13 \pm 0.35	3.88 \pm 0.35	3.00 \pm 0.00
Structure preservation	1.5 \pm 0.53	1.75 \pm 0.46	3.12 \pm 0.83	3.38 \pm 0.52	3.75 \pm 0.46	3.75 \pm 0.46	3.75 \pm 0.46	4.00 \pm 0.00
Continuity between slices	2.25 \pm 0.71	2.38 \pm 0.74	1.88 \pm 1.13	2.63 \pm 0.92	3.50 \pm 1.07	3.75 \pm 0.46	3.88 \pm 0.35	3.75 \pm 0.46
Sharpness	1.63 \pm 0.52	1.75 \pm 0.46	2.88 \pm 0.64	3.25 \pm 0.71	3.75 \pm 0.46	3.88 \pm 0.35	3.75 \pm 0.46	3.25 \pm 0.46
Artifact reduction	3.00 \pm 0.00	3.00 \pm 0.00	2.50 \pm 1.07	2.63 \pm 0.92	3.63 \pm 0.74	3.75 \pm 0.46	3.75 \pm 0.46	3.88 \pm 0.53

**Fig. 8** Comparison of MSE loss values versus the number of epochs with respect to different residual CNN models.

according to the score of overall image quality, but the difference was not statistically significant compared with either ResNet2D_7Slices or ResNet3D-v1 ($p > 0.05$).

We calculated the averaged MSE loss value achieved by different methods versus the number of epochs as the measure of convergence in Fig. 8. From this figure, we observe that all models have similar convergence trends in that all validation loss curves decreased initially and then converged around 35th epoch. When the input channels of the 2D CNN model increased from one to three to seven, a trend of a lower MSE loss was observed. The ResNet3D-v1 model achieved a lower MSE loss than the ResNet2D_7Slices model. The lowest MSE loss was achieved by the ResNet3D-v2 model, although it was trained for only 38 epochs (totally around 171 h).

4 Discussion and Conclusions

In this work, we investigated the impact of multislice input in deep CNN CT denoising on image quality by comparing 2D residual CNN models with different numbers of input slices, and

among the 2D and two versions of 3D residual CNN models with the seven-slice input. To the best of our knowledge, this study was the first to evaluate if a multislice input improves the denoising performance compared with a single-slice input and if a 3D network architecture is better than a 2D version at utilizing the multislice input.

The evaluation results clearly demonstrated that the denoising performance was improved by including more CT slices as the input of 2D residual CNN.

With the same number of CT slices as the input, two versions of 3D residual CNNs, especially the fully 3D model (ResNet3D-v2), seem to be superior to the corresponding 2D model in terms of artifacts and continuity between slices. However, the overall image training time was significantly longer for the fully 3D model. The performance of 2D and two versions of 3D residual CNN models with the seven-slice input may be further improved with more slices as the input, but the number of slices should be carefully selected for various applications in terms of slice thickness. The relationship and redundancy that can be leveraged between slices persists but eventually vanishes due to distance.

The 2D residual CNN models using a stack of 2D slices as input were referred to as 2.5D CNN in some prior publications.^{25–27} This 2.5D CNN can exploit 3D geometric information while still using a 2D CNN architecture. Although some researchers stated that the 3D structures may be lost by the first 2D convolution in the 2.5D CNN model,²⁸ others believed that the 2.5D CNN can often achieve an image quality similar to 3D CNNs at a reduced computational cost.^{26,27,29} In this work, we did not refer to this model as “2.5D CNN” because this name was also used to represent a CNN model that uses both 2D and 3D convolutions in the same model,³⁰ which is similar to our ResNet3D-v1 model.

We investigated two versions of 3D CNN models for comparison, as shown in Figs. 2 and 3, respectively. The architecture of ResNet3D-v2 model is similar to that in Ref. 18, except that the sizes of the training patches were different ($64 \times 64 \times 7$ in our study versus $44 \times 44 \times 24$ in the previous work). The ResNet3D-v1 model was designed following the same idea in Ref. 20, which used a 3D input and a 2D output, except that we used a residual net instead of Unet. Despite those similarities, neither study in Refs. 18 and 20 included 2D CNN models with multiple slices as the input to compare with their proposed 3D CNN models. Although both ResNet3D-v1 and ResNet3D-v2 require more processing than the 2D residual CNN models with the seven-slice input, the former is less computationally expensive than the latter: 3D convolution layers are used in the first half and the entire model for ResNet3D-v1 and ResNet3D-v2, respectively. Pure 3D CNN models are more popular in the research field of anatomical segmentation.²⁵ In this study, the pure 3D ResNet3D-v2 model delivers better performance than the ResNet3D-v1 model based on the radiologist’s subjective evaluation.

There are several limitations in this research such as the relatively small sample size used in the analysis. In future research, more cases will be included for nonparametric statistical analysis and to determine the reliability of the statistical findings. Additionally, our study used only one radiologist. We will use more radiologists for visual assessment of the images to exclude bias arising from an observer variation.

We conclude that the multislice input is an effective strategy for improving the performance for 2D deep CNN denoising models. Pure 3D CNN model tends to have a better performance than the other models in terms of continuity across axial slices, but the difference was not significant compared with the 2D CNN model with the same number of slices as the input.

Disclosures

No potential conflicts of interest were declared.

Acknowledgments

The research reported in this work was supported by the Mayo Clinic Department of Radiology Scholarship program and the CT Clinical Innovation Center. Some content was published in the SPIE Medical Imaging Conference proceeding in 2022. The current manuscript has some

changes and new content from the previous proceeding paper. Institutional review board approval was acquired to use CT projection and image data from patient ablation procedures.

References

1. B. Newman and M. J. Callahan, "ALARA (as low as reasonably achievable) CT 2011—executive summary," *Pediatr. Radiol.* **41**(Suppl 2), 453–455 (2011).
2. M. Balda, J. Hornegger, and B. Heismann, "Ray contribution masks for structure adaptive sinogram filtering," *IEEE Trans. Med. Imaging* **31**(6), 1228–1239 (2012).
3. A. Manduca et al., "Projection space denoising with bilateral filtering and CT noise modeling for dose reduction in CT," *Med. Phys.* **36**(11), 4911–4919 (2009).
4. E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Phys. Med. Biol.* **53**(17), 4777–4807 (2008).
5. Y. Zhang et al., "Statistical iterative reconstruction using adaptive fractional order regularization," *Biomed. Opt. Express* **7**(3), 1015–1029 (2016).
6. Y. Chen et al., "Bayesian statistical reconstruction for low-dose X-ray computed tomography using an adaptive-weighting nonlocal prior," *Comput. Med. Imaging Graphics* **33**(7), 495–500 (2009).
7. S. J. Chang et al., "Spectrum estimation-guided iterative reconstruction algorithm for dual energy CT," *IEEE Trans. Med. Imaging* **39**(1), 246–258 (2020).
8. Z. Li et al., "Adaptive nonlocal means filtering based on local noise level for CT denoising," *Med. Phys.* **41**(1), 011908 (2014).
9. M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.* **54**(11), 4311–4322 (2006).
10. K. Sheng et al., "Denoised and texture enhanced MVCT to improve soft tissue conspicuity," *Med. Phys.* **41**(10), 101916 (2014).
11. A. Mileto et al., "State of the art in abdominal CT: the limits of iterative reconstruction algorithms," *Radiology* **293**(3), 491–503 (2019).
12. K. Li, J. Tang, and G.H. Chen, "Statistical model based iterative reconstruction (MBIR) in clinical CT systems: experimental assessment of noise performance," *Med Phys* **41**(4), 041906 (2014).
13. H. Chen et al., "Low-dose CT via convolutional neural network," *Biomed. Opt. Express* **8**(2), 679–694 (2017).
14. Q. S. Yang et al., "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).
15. H. Chen et al., "Low-dose CT with a residual encoder-decoder convolutional neural network," *IEEE Trans. Med. Imaging* **36**(12), 2524–2535 (2017).
16. N. R. Huber et al., "Evaluating a convolutional neural network noise reduction method when applied to CT images reconstructed differently than training data," *J. Comput. Assist. Tomogr.* **45**(4), 544–551 (2021).
17. M. Li et al., "SACNN: self-attention convolutional neural network for low-dose CT denoising with self-supervised perceptual loss network," *IEEE Trans. Med. Imaging* **39**(7), 2289–2301 (2020).
18. W. Yang et al., "Improving low-dose CT image using residual convolutional network," *IEEE Access* **5**, 24698–24705 (2017).
19. C. Y. You et al., "Structurally-sensitive multi-scale deep neural network for low-dose CT denoising," *IEEE Access* **6**, 41839–41855 (2018).
20. H.M. Shan et al., "3-D convolutional encoder-decoder network for low-dose CT via transfer learning from a 2-D trained network," *IEEE Trans. Med. Imaging* **37**(6), 1522–1534 (2018).
21. Z. Zhou et al., "Residual-based convolutional-neural-network (CNN) for low-dose CT denoising: impact of multi-slice input," *Proc. SPIE* **12031**, 120312B (2022).

22. C. H. McCollough et al., “Low-dose CT for the detection and classification of metastatic liver lesions: results of the 2016 low dose CT grand challenge,” *Med. Phys.* **44**(10), e339–e352 (2017).
23. D. Kingma and J. Ba, “Adam: a method for stochastic optimization,” in *Int. Conf. Learn. Represent., ICLR ('15)* (2015).
24. C.T. Jensen et al., “Image quality assessment of abdominal CT by use of new deep learning image reconstruction: initial experience,” *Am. J. Roentgenol.* **215**(1), 50–57 (2020).
25. J. Bernal et al., “Deep convolutional neural networks for brain image analysis on magnetic resonance imaging: a review,” *Artif. Intell. Med.* **95**, 64–81 (2019).
26. A. Ziabari et al., “2.5D deep learning for CT image reconstruction using a multi-GPU implementation,” in *52nd Asilomar Conf. Signals, Syst. and Comput.*, pp. 2044–2049 (2018).
27. U. Javaid et al., “Mitigating inherent noise in Monte Carlo dose distributions using dilated U-Net,” *Med. Phys.* **46**(12), 5790–5798 (2019).
28. H. Zheng et al., “Improving the slice interaction of 2.5D CNN for automatic pancreas segmentation,” *Med. Phys.* **47**(11), 5543–5554 (2020).
29. A. A. Hendriksen et al., “Deep denoising for multi-dimensional synchrotron X-ray tomography without high-quality reference data,” *Sci. Rep.* **11**(1), 11895 (2021).
30. Y. Xue et al., “A multi-path 2.5 dimensional convolutional neural network system for segmenting stroke lesions in brain MRI images,” *Neuroimage Clin.* **25**, 102118 (2020).

Zhongxing Zhou received his bachelor’s, master’s, and PhD degrees in biomedical engineering in 2003, 2006, and 2009, all from Tianjin University, Tianjin, China, respectively. He is a senior research fellow at Mayo Clinic. His current research interests include low-dose CT imaging techniques, CT image quality assessment, and deep learning-based CT imaging.

Nathan R. Huber received his BA degree in physics and chemistry from Gustavus Adolphus College in 2017 and his PhD in biomedical engineering and physiology from Mayo Clinic Graduate School in 2022. He is a CT systems scientist at GE HealthCare. His research interests include deep learning image processing and photon counting CT.

Akitoshi Inoue received his MD degree in 2007 and his PhD in radiology in 2017 both from Shiga University of Medical Science. He is an assistant professor of radiology at Mayo Clinic and Shiga of the University of Medical Science. He works as a radiologist in Japan and collaborates with Mayo Clinic for research. His research interest is clinical validation of CT technology and gastrointestinal imaging, such as gastrointestinal stromal tumor and rectal cancer.

Cynthia H. McCollough received her doctorate degree from the University of Wisconsin in 1991. She is a professor of the medical physics and biomedical engineering at Mayo Clinic, where she directs the CT Clinical Innovation Center. Her research interests include CT dosimetry, advanced CT technology, and innovative clinical applications, such as dual-energy and multispectral CT. She is an NIH-funded investigator and is active in numerous professional organizations. He is a fellow of the AAPM, ACR, and AIMBE.

Lifeng Yu received his BS degree in nuclear physics in 1997 and his MEng degree in nuclear technology in 2000, both from Beijing University, and his PhD in medical physics from the University of Chicago in 2006. He is a professor of medical physics at Mayo Clinic and a fellow of the AAPM and SPIE. His research interests include CT physics, image quality assessment, radiation dose reduction, and spectral CT.