Original Article

Transferring U-Net between low-dose CT denoising tasks: a validation study with varied spatial resolutions

Xin Zhang[1,2], Ting Su[1], Yunxin Zhang[3], Han Cui[1], Yuhang Tan[1], Jiongtao Zhu[4], Dongmei Xia[5], Hairong Zheng[6], Dong Liang[1,6], Yongshuai Ge[1,6]

[1]Research Center for Medical Artificial Intelligence, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China; [2]University of Chinese Academy of Sciences, Beijing, China; [3]Department of Vascular Surgery, Beijing Jishuitan Hospital, Beijing, China; [4]College of Physics and Optoelectronic Engineering, Shenzhen University, Shenzhen, China; [5]Key Laboratory of Low-grade Energy Utilization Technologies and Systems of Ministry of Education of China, College of Power Engineering, Chongqing University, Chongqing, China; [6]Paul C Lauterbur Research Center for Biomedical Imaging, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China

Contributions: (I) Conception and design: X Zhang, T Su, D Xia, Y Ge; (II) Administrative support: H Zheng, D Liang, Y Ge; (III) Provision of study materials or patients: X Zhang, T Su, Y Ge; (IV) Collection and assembly of data: X Zhang, T Su, Y Ge; (V) Data analysis and interpretation: X Zhang, T Su, Y Zhang, J Zhu, Y Tan, H Cui, D Xia, Y Ge; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Dongmei Xia, PhD. Key Laboratory of Low-grade Energy Utilization Technologies and Systems of Ministry of Education of China, College of Power Engineering, Chongqing University, 174 Shazheng Street, Shapingba District, Chongqing 400044, China. Email: xiadm@cqu.edu.cn; Dong Liang, PhD; Yongshuai Ge, PhD. Research Center for Medical Artificial Intelligence, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, 1068 Xueyuan Avenue, Nanshan District, Shenzhen 518055, China; Paul C Lauterbur Research Center for Biomedical Imaging, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China. Email: dong.liang@siat.ac.cn; ys.ge@siat.ac.cn.

Background: Recently, deep learning techniques have been widely used in low-dose computed tomography (LDCT) imaging applications for quickly generating high quality computed tomography (CT) images at lower radiation dose levels. The purpose of this study is to validate the reproducibility of the denoising performance of a given network that has been trained in advance across varied LDCT image datasets that are acquired from different imaging systems with different spatial resolutions.

Methods: Specifically, LDCT images with comparable noise levels but having different spatial resolutions were prepared to train the U-Net. The number of CT images used for the network training, validation and test was 2,400, 300 and 300, respectively. Afterwards, self- and cross-validations among six selected spatial resolutions (62.5, 125, 250, 375, 500, 625 μm) were studied and compared side by side. The residual variance, peak signal to noise ratio (PSNR), normalized root mean square error (NRMSE) and structural similarity (SSIM) were measured and compared. In addition, network retraining on a small number of image set was performed to fine tune the performance of transfer learning among LDCT tasks with varied spatial resolutions.

Results: Results demonstrated that the U-Net trained upon LDCT images having a certain spatial resolution can effectively reduce the noise of the other LDCT images having different spatial resolutions. Regardless, results showed that image artifacts would be generated during the above cross validations. For instance, noticeable residual artifacts were presented at the margin and central areas of the object as the resolution inconsistency increased. The retraining results showed that the artifacts caused by the resolution mismatch can be greatly reduced by utilizing about only 20% of the original training data size. This quantitative improvement led to a reduction in the NRMSE from 0.1898 to 0.1263 and an increase in the SSIM from 0.7558 to 0.8036.

Conclusions: In conclusion, artifacts would be generated when transferring the U-Net to a LDCT denoising task with different spatial resolution. To maintain the denoising performance, it is recommended to retrain the U-Net with a small amount of datasets having the same target spatial resolution.

Keywords: Low-dose computed tomography (LDCT); computed tomography (CT) denoising; spatial resolution; network retraining; U-Net

Submitted May 30, 2023. Accepted for publication Nov 09, 2023. Published online Jan 02, 2024.
doi: 10.21037/qims-23-768
View this article at: https://dx.doi.org/10.21037/qims-23-768

Introduction

The recent booming of deep learning (DL) techniques has substantially shifted the paradigm of low-dose computed tomography (LDCT) medical imaging, which aims at reducing the radiation dose of CT scans to as low as reasonably achievable (ALARA). Many studies (1-10) have investigated to seek for the advanced LDCT imaging solutions, especially for the way of generating and reconstructing CT images with higher signal-to-noise ratio (SNR) at lower radiation dose. Different from the well-known filtered back-projection (FBP) method and the iterative reconstruction (IR) method, the performance of DL approaches heavily relies on the training data, which is usually made by two sets of CT images: the expected images, i.e., the label images obtained at standard exposures; the images to be processed, i.e., the input CT images obtained at lower exposures. Afterwards, the training procedure makes the network learn to remove the noise on the LDCT images.

As required by the DL method, dedicated LDCT data always have to be gathered and prepared in advance. Take the worldwide recognized American Association of Physicists in Medicine (AAPM) LDCT imaging challenge data (11) as an example, those LDCT images were prepared for low-mAs CT imaging applications with conventional imaging spatial resolution, i.e., $\Delta x > 625$ μm. In this study, Δx denoted the spatial resolution of the imaging system by default. When dealing with a different LDCT imaging task of varied spatial resolution (12), for instance, the small animal biological micro-CT imaging (13) with a spatial resolution of $\Delta x < 50$ μm, it is hard to find such a well-recognized and open-sourced LDCT imaging data. Intuitively, it is viable to use the AAPM clinical LDCT image data in the low-dose micro-CT imaging applications (*Figure 1*). Herein, the similar object contents were assumed, namely, the anatomic structures of small animals were assumed to be similar to that of humans. Essentially, this leads to a very interesting question: can the neural network trained at one spatial resolution be transferred and applied directly onto another LDCT imaging task with different spatial resolution, provided that both the noise level and structural content are similar? If the answer to this question is positive, a specific LDCT imaging network can be easily extended without spending a lot of resources to retrain it. Consequently, many efforts such as the data preparation, network training, computation and human resources can be saved. To answer the above question, the LDCT imaging performance of a popular DL algorithm, e.g., U-Net (14), at six varied image spatial resolutions is validated in this study.

In the field of DL, some preprocessing is usually necessary if a pre-trained network model needs to be applied onto other tasks. And the most common method is transfer learning (15-18), which has been proven effective in improving the performance of neural networks on various tasks. Transfer learning can be especially useful in situations where a limited amount of data is available for the target task or when the new task is closely related to the original task. The application of the varied resolution network model in this study also involves limitations in data quantity and similarity of tasks. Therefore, some transfer learning processes have been applied to relevant network models in this study. Without changing the network structure, six sets of small amounts of target resolution image data were used to retrain the non-target resolution image model, and then applied to the denoising task of target resolution LDCT images. Through such transfer learning research, it is possible to investigate whether the generalization ability of network can be improved after the network retraining which was based on a small amount of target task data.

The rest of this article is organized as follows: the Methods section presents the CT imaging physics, numerical experiments, and details of the network training and retraining; the Results section presents the LDCT imaging results at six certain image resolutions and the network retraining results; the Discussion and Conclusions section present the discussions and a brief conclusion of this study.

Methods

CT imaging model

In CT imaging, the measured beam intensity I in detector

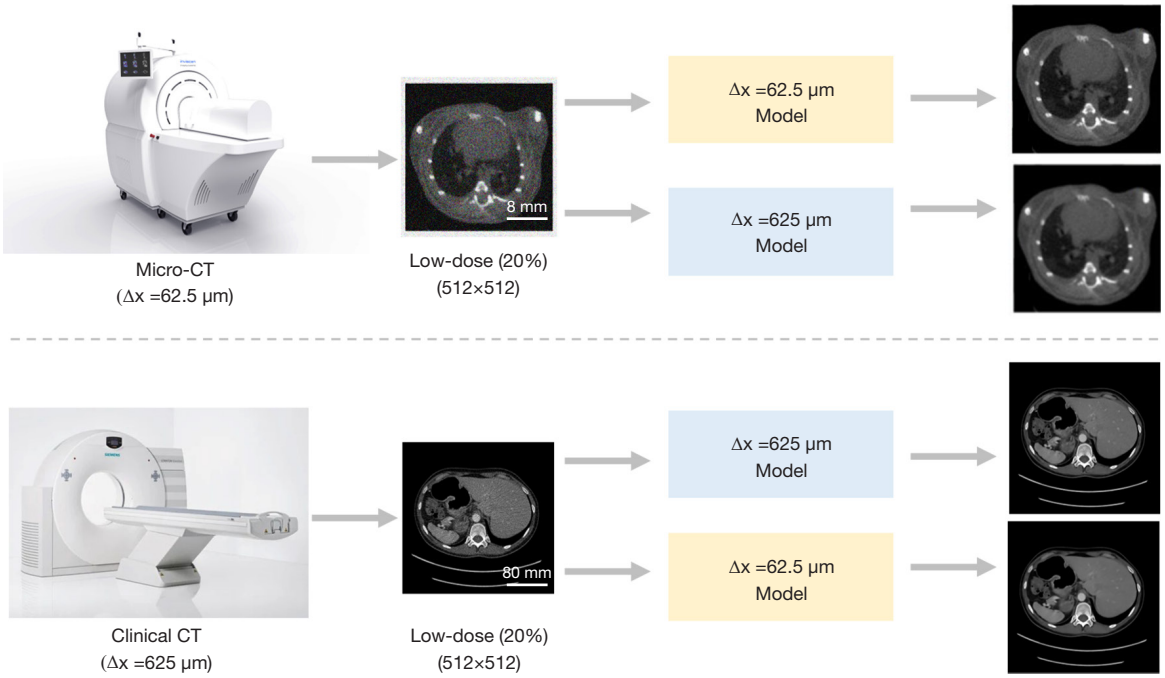

Figure 1 Illustrations of the LDCT imaging tasks at two spatial resolution levels (Δx =62.5 μm and Δx =625 μm). LDCT, low-dose computed tomography.

element t at projection angle θ is expressed as

$$I(t,\theta)=I_{0,A}\int\Omega(E)\exp\left(-\mathcal{R}_{\mu}(t,\theta,E)\right)dE \quad [1]$$

where $I_{0,A}$ denotes the number of incident X-ray photons, $\Omega(E)$ denotes the X-ray beam spectra, and $\mathcal{R}_{\mu}(\cdot)$ denotes the Radon transform (19) of the object $\mu(\vec{x},E)$:

$$\mathcal{R}_{\mu}(t,\theta,E)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mu(\vec{x},E)\tilde{\delta}(\vec{x}\cdot\vec{n}_{\theta}-t)\,dxdy \quad [2]$$

in which $\tilde{\delta}(\cdot)$ denotes the Dirac δ-function and represents an X-ray beamlet, $\vec{x}=(x,y)$, and $\vec{n}_{\theta}=(\cos\theta,\sin\theta)$. At last, CT images are reconstructed using the FBP algorithm along with Ramp filter.

Data preparation

In this study, a group of selected abdominal CT images were downloaded from the Cancer Imaging Archive (20) at https://www.cancerimagingarchive.net/ for network training and validation. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). Before data preparations, these CT images having Hounsfield unit (HU) were converted into linear attenuation coefficient via formula: $\mu_{object}=\frac{HU}{1000}\times\mu_{water}+\mu_{water}$. Herein, the X-ray energy was assumed as 60 keV, and the μ_{water} at 60 keV was obtained from National Institute of Standards and Technology (NIST, USA).

Forward projection and FBP reconstructions were performed to generate the needed noisy CT images. In total, six spatial imaging resolutions (denoting the resolving capability of the image system) were manually selected: 62.5, 125, 250, 375, 500 and 625 μm. The main parameters are listed in *Table 1*. It should be noted that these resolution settings were not corresponding to certain imaging modalities, and each setting only represented a typical class of imaging modality. For example, the 62.5 μm setting denoted the typical spatial resolution of a small animal micro-CT scanner, and the 625 μm setting denoted the typical spatial resolution of a clinical CT scanner. With Poisson statistics, the used photons enable us to generate CT images with similar SNRs. No electronic noise was added in these numerical experiments. The X-ray source to detector distance (SDD) and the source to rotation center (SOD) were adjusted adaptively along with the resolution settings. Meanwhile, the dimension of the detector, denoted as Δ_{det}, also varied along with Δ_x. By default, $\Delta_{det}=\frac{SDD}{SOD}\times\Delta x$ was assumed in this work. For more

Table 1 The main configurations of the six numerical CT imaging systems

Spatial resolution	62.5 μm	125 μm	250 μm	375 μm	500 μm	625 μm
Input I_0 (photons)	3.6×10^4	1.5×10^4	1×10^4	1.5×10^4	2.5×10^4	5.6×10^4
Label I_0 (photons)	1.8×10^5	7.5×10^4	5×10^4	7.5×10^4	1.25×10^5	2.8×10^5
SDD (mm)	150	300	600	900	1,200	1,500
SOD (mm)	75	150	300	450	600	750

The input and label photon numbers (I_0) are listed in the first and second rows, the X-ray SDD and the SOD are listed in the third and fourth rows. CT, computed tomography; SDD, source to detector distance; SOD, source to rotation center.

details, please refer to our previous work (6,21). The animal experiment was approved by the Institutional Animal Care and Use Committee (IACUC) of the Shenzhen Institute of Advanced Technology at the Chinese Academy of Sciences and was conducted in compliance with the protocol (SIAT-IACUC-201228-YGS-LXJ-A1498; January 5, 2021) for the care and use of animals.

The U-Net

In this study, the U-Net (14) proposed by Ronneberger *et al.* was trained. Specifically, there are 24 convolutional layers in total. The contracting branch contains 11 convolutional layers, and the number of feature channels in these layers is 32, 32, 32, 64, 64, 128, 128, 256, 256, 512 and 512, respectively. The size of the convolutional kernel is 3×3 for each layer in contracting branch, and 2×2 max pooling operations with stride 2 are applied for down-sampling. The expansive branch contains 13 convolutional layers, and the number of feature channels in these layers is 256, 256, 256, 128, 128, 128, 64, 64, 64, 32, 32, 32 and 1, respectively. The size of the convolutional kernel is 3×3 for the first layer to the twelfth layer in expansive branch, and the size of the convolutional kernel is 1×1 for the final layer in expansive branch. Additionally, 3×3 deconvolution operations with stride 2 are applied for up-sampling. The activation function was replaced by the leaky rectified linear unit (Leaky ReLU). Concatenations were added between the corresponding layers having the same image size. In addition, the mean-square-error (MSE) is selected as the network loss, which is defined as:

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left[y(i,j)-\hat{y}(i,j)\right]^2 \quad [3]$$

where y denotes the label CT images with normal dose, $\hat{y}$ denotes the predicted image from the network, and $m=n=512$ denote the size of the images. All network training were implemented in Python with the TensorFlow library on a NVIDIA RTX A6000 GPU card.

Network training

In total, six U-Net models were trained with respect to the LDCT images having six different spatial resolutions. For a certain image resolution, 2,400 images were used for the U-Net training, 300 images were used for validation and 300 images were used for testing. Finally, the LDCT images with high resolution (Δx =62.5 μm) and the LDCT images with low resolution (Δx =625 μm) are validated by all these six network models, correspondingly. The peak signal to noise ratio (PSNR), normalized root mean square error (NRMSE) and structural similarity (SSIM) are measured to evaluate the denoising performance. The PSNR is defined as,

$$PSNR = 10\cdot\log_{10}\left(\frac{MAX_I^2}{MSE}\right) \quad [4]$$

where MAX_I denotes the maximum intensity value, and the MSE used here is calculated based on the ground truth. And the NRMSE can be denoted as

$$NRMSE = \frac{\sqrt{MSE}}{\bar{y}}\times100 \quad [5]$$

where $\bar{y}$ is the mean of the cross-test predicted image, and the MSE used here is calculated based on the corresponding self-test result. In addition, the SSIM is used to measure the perceptual similarity between the cross-test LDCT and the self-test CT images, which is defined as

$$SSIM(\hat{y},y) = \frac{(2\mu_{\hat{y}}\mu_y + c_1)(2\sigma_{\hat{y}y} + c_2)}{(\mu_{\hat{y}}^2 + \mu_y^2 + c_1)(\sigma_{\hat{y}}^2 + \sigma_y^2 + c_2)} \quad [6]$$

where $\hat{y}$ denotes the self-test image, y denotes the cross-test image. And $\mu_{\hat{y}}$ and μ_y denote the averages of $\hat{y}$ and y, $\sigma_{\hat{y}}^2$ and σ_y^2 denote the variance of $\hat{y}$ and y, $\sigma_{\hat{y}y}$ denotes the

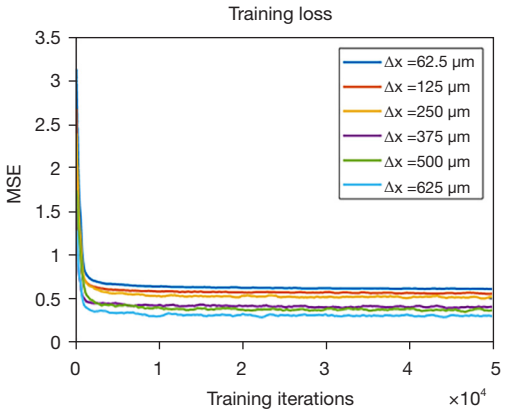

Figure 2 The U-Net training loss curves with six spatial resolutions CT images. The MSE is selected as the network loss function. MSE, mean-square-error; CT, computed tomography.

Table 2 The training time used by each model

Model	Training steps	Computation time
62.5 μm	50,000	6 h 48 min
125 μm	50,000	6 h 52 min
250 μm	50,000	6 h 48 min
375 μm	50,000	6 h 47 min
500 μm	50,000	6 h 49 min
625 μm	50,000	6 h 48 min

co-variance between $\hat{y}$ and y, and c_1 and c_2 denote the two variables to stabilize the division operation.

Model retraining

Additionally, network retraining was performed for the 625 μm model by a small amount of LDCT images with 62.5 μm spatial resolution. To do so, six different amounts of data were utilized, corresponding to 5%, 10%, 15%, 20%, 25% and 30% of the original 2,400 training images. The number of images used in both validation and testing sets was fixed at 300. Afterwards, such retrained model was used to test the 62.5 μm LDCT images. Early stopping, namely, stopping the network training if the loss value does not decrease after a certain number of epochs, is used during the network retraining. Specifically, the number of retraining steps for the six groups was: 3,000 for 5% group, 5,000 for 10%, 15% and 20% groups, 6,000 for 25% group, and 7,000 for 30% group.

Results

Network loss

The training losses are plotted in *Figure 2*. It can be seen that the loss curves of six different resolutions exhibit similar trends, indicating the independence of the U-Net training to the spatial resolution. These CT images of different resolutions used for network training were generated with similar SNRs. The presented figure illustrates that despite the network model trained on high spatial resolution images exhibiting higher MSE values than the model trained on low resolution images, all training curves converge within an equal number of training steps. The final convergent values of these training loss curves in *Figure 2* are different. We guess this might be due to the slightly different noise levels of the generated CT images at different spatial resolutions. Additionally, the network training time is listed in *Table 2*. Approximately, the total network training time is quite close from each other.

Results of high-resolution images

The high-resolution LDCT images (62.5 μm) were denoised by the U-Net model trained at the same resolution (denoted as self-test) and the U-Net model trained at other resolutions (denoted as cross-test). Results in *Figure 3* demonstrate that the trained network can be used to denoise the LDCT images that have the same resolution without generating significant image artifacts. Herein, two different abdominal structures were compared. Results on *Figure 4* are for the cross-tests with the other LDCT images having varied spatial resolutions. Two region-of-interests (ROIs), highlighted by the yellow and blue boxes, were selected. In addition, the difference maps with respect to the self-test result are depicted in the second and fourth rows. Obviously, these U-Net trained at different image spatial resolutions can be used to remove the image noise on the high resolution LDCT images. The PSNR values listed in *Table 3* also indicate that all models can denoise the LDCT images having 62.5 μm spatial resolution.

Aside from the noise reduction, however, it can be noticed that some fake image contents were generated during the cross-test validations. For the tested high resolution LDCT images, more artificial textures show up at the edges of the images as the resolution used in U-Net model varies from 125 μm up to 625 μm, see the yellow ROIs on *Figure 4*. On the contrary, such artifacts are not significant for the blue ROIs located around the central

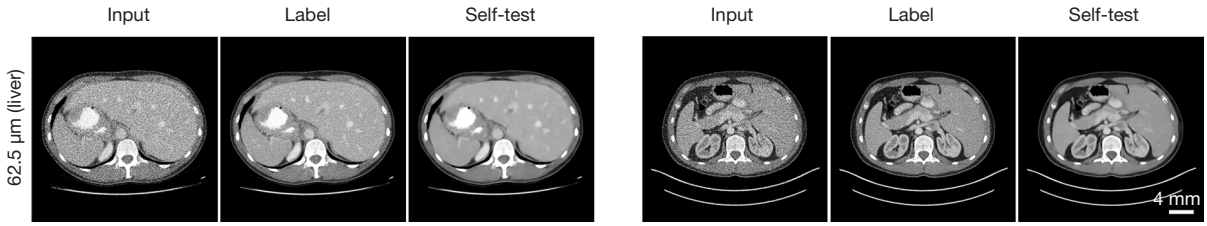

Figure 3 The U-Net self-test results of 62.5 μm CT images (including the images of liver and pancreas). The display window is [0.17, 0.26] cm^{-1}. The scale-bars denote 4.0 mm. CT, computed tomography.

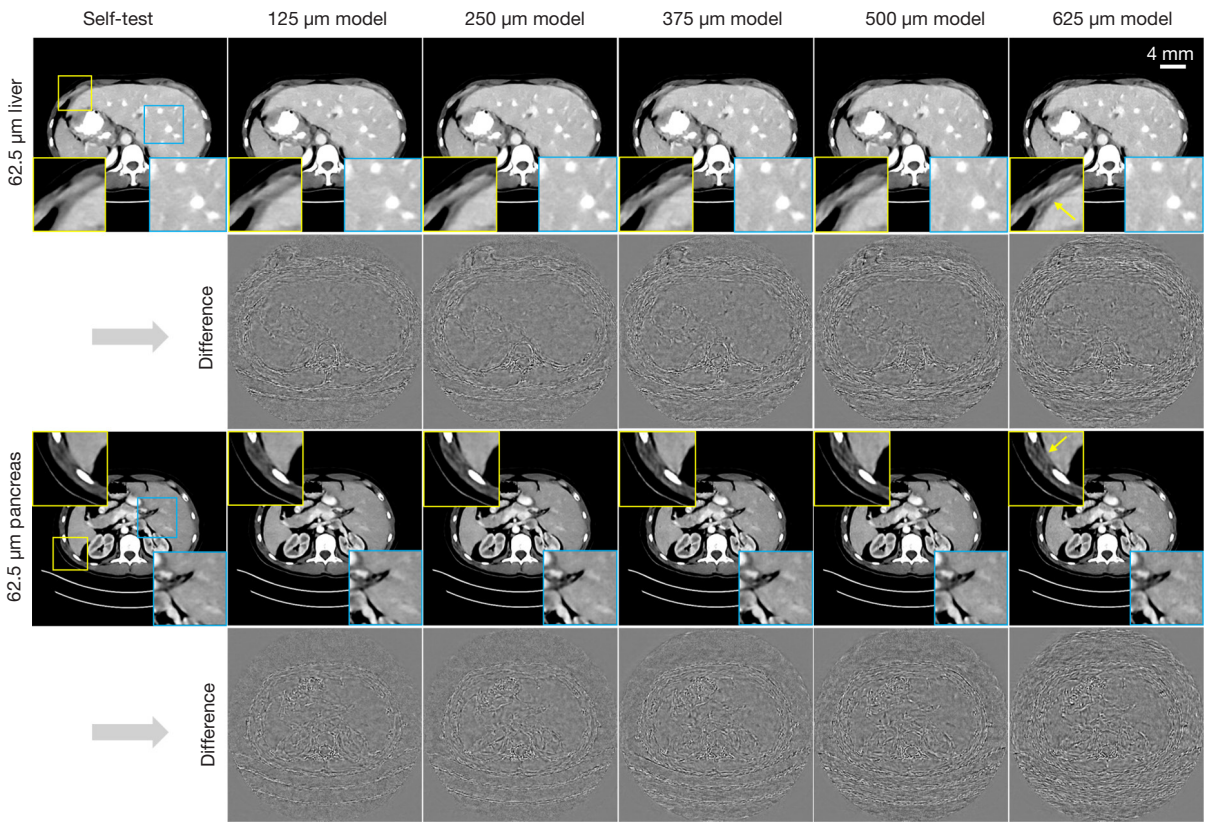

Figure 4 The U-Net test results of 62.5 μm CT images. The network test images are shown in the first and third rows, and the corresponding residual images which obtained by subtracting the self-test image from the cross-test images are shown in the second and fourth rows. The display window of test images is [0.20, 0.24] cm^{-1} and the display window of residual images is [–0.007, 0.007] cm^{-1}. The scale-bars denote 4.0 mm. CT, computed tomography.

vicinity. In addition, the number of residual artifacts become dramatic as the resolution discrepancy increases between the training model and the targeting validation model, see the difference images on *Figure 4*.

More quantitative analysis results are presented in *Figure 5*. In particular, the variance of the residual images is compared on *Figure 5A*, NRMSEs and SSIMs of the validation results are compared on *Figure 5B*,*5C*, respectively. In short, these quantitative results show high consistency with the above visual observations. As the resolution decreases, the residual variance and NRMSE values get larger, and the SSIM values get smaller, indicating a worse performance in maintaining the identical image information. In other words, cross tests with unmatched

Table 3 The measured PSNR values (dB) of 62.5 μm and 625 μm test results

Test	PSNR (dB)					
	62.5 μm model	125 μm model	250 μm model	375 μm model	500 μm model	625 μm model
62.5 μm-LDCT	37.17±1.03	37.26±1.03	37.31±1.03	37.11±1.01	36.98±1.01	36.87±1.00
625 μm-LDCT	37.06±1.01	37.43±1.04	37.71±1.05	37.73±1.06	37.71±1.06	37.75±1.06

The values are presented as mean ± standard deviation. LDCT, low-dose computed tomography; PSNR, peak signal to noise ratio.

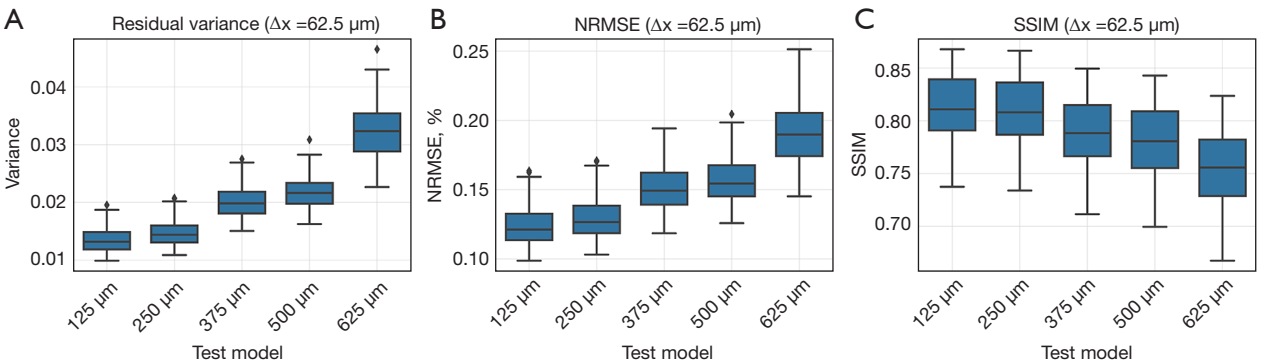

Figure 5 The quantitative results of the cross-test results of 62.5 μm images. (A) The variance values of residual images; (B) NRMSEs of the test images; (C) SSIMs of the test images. NRMSEs, normalized root mean square errors; SSIMs, structural similarities.

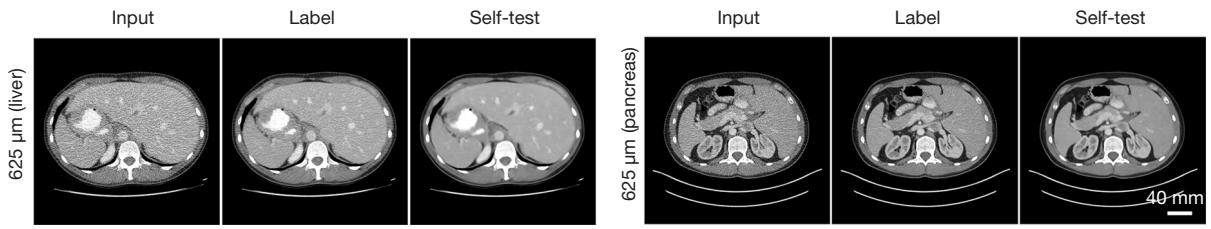

Figure 6 The U-Net self-test results of 625 μm CT images (including the images of liver and pancreas). The display window is [0.17, 0.26] cm^{-1}. The scale-bars denote 40.0 mm. CT, computed tomography.

spatial resolution may introduce alien image information, e.g., random image artifacts, and thus degrades the overall image quality.

Results of low-resolution images

This study also explored the low-resolution (625 μm) LDCT imaging performance in a similar manner. *Figure 6* depicts the self-test results. Likewise, the outcomes indicate that the trained network can be employed to effectively denoise the LDCT images having the same resolution. Herein, two different abdominal structures were compared. The cross-test results presented in *Figure 7* demonstrate that the U-Net models trained with other resolution datasets can be used to denoise the low-resolution (625 μm) LDCT images. This is consistent with the results obtained from the validations of high-resolution (62.5 μm) LDCT data. Specifically, two ROIs were highlighted and the corresponding residual images are illustrated in the second and fourth rows. *Table 3* presents the quantitative analysis results, which also demonstrate that all the U-Net models are capable of effectively denoise the low-resolution LDCT images.

In contrast to the high-resolution validations, the

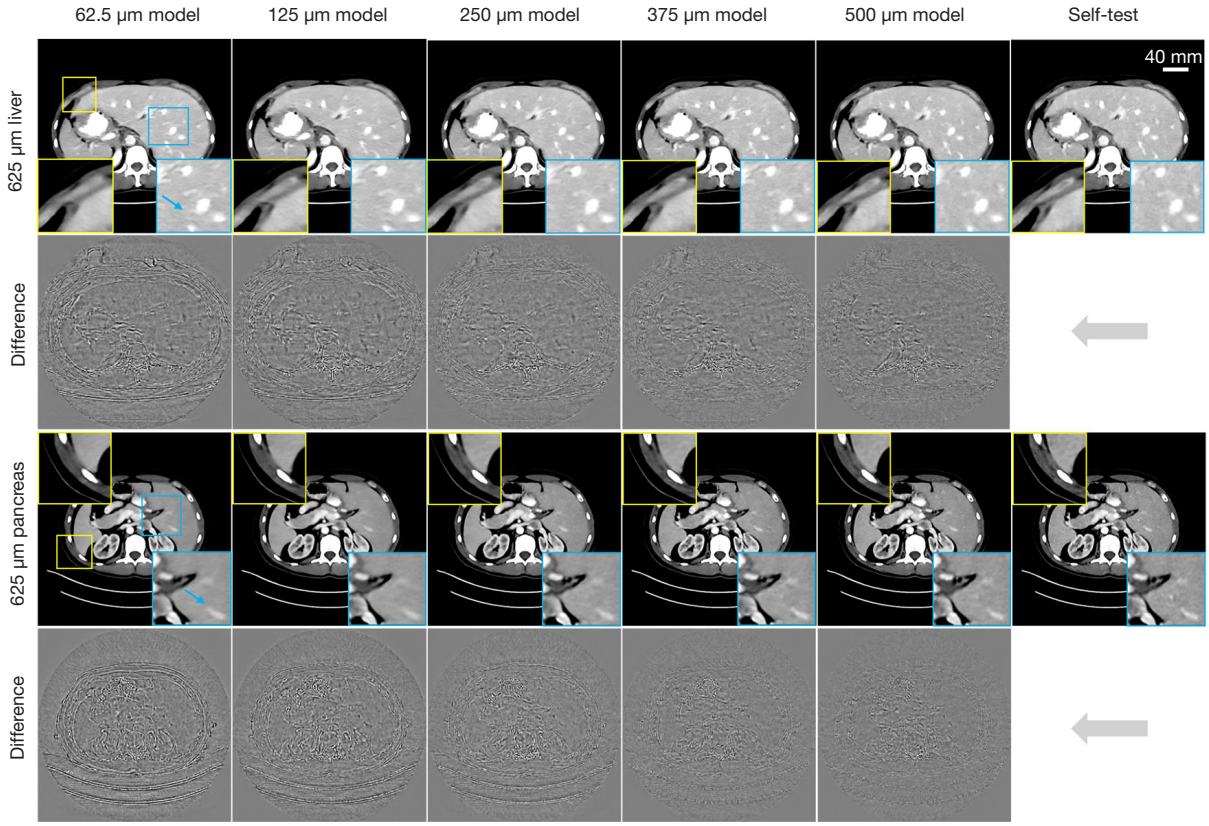

Figure 7 The U-Net test results of 625 μm CT images. The layout of these images is similar to *Figure 4*. The display window of test images is [0.20, 0.24] cm^{-1} and the display window of residual images is [–0.007, 0.007] cm^{-1}. The scale-bars denote 40.0 mm. CT, computed tomography.

occurrence of artifacts on the periphery of the CT images was rare and less prominent in such low-resolution validations, see the yellow ROIs in *Figure* 7. However, the sharpness of the denoised low-resolution LDCT images gradually degrades as the spatial resolution of the training data increases from 500 to 62.5 μm, see the two highlighted ROIs on *Figure* 7. As the resolution disparity between the training data and testing data increases, the amount of residual artifacts also escalates, see the residual images in *Figure* 7.

The measured residual variance, NRMSEs, and SSIMs results were plotted in *Figure 8*, respectively. These quantitative results indicate a gradual degradation of the cross-test outcomes as the spatial resolution varies. These findings again demonstrate that the network validation using inconsistent resolutions would introduce artifacts to the denoised LDCT images, thereby diminishing the generalization capability among different network models.

Results of network retraining

With the purpose to enhance the LDCT denoising performance of the cross tests, network retraining was investigated with the low-resolution model (625 μm). Namely, the U-Net that has already been trained by the low-resolution LDCT images was retrained by a small fraction of high-resolution LDCT images (62.5 μm). The retraining losses are plotted in *Figure 9*. It should be noted that each group utilized different training data, and to avoid overfitting, the number of training steps performed by each group varied. Therefore, each group stopped training at the point of their lowest loss value, which was approximately similar across the groups.

Afterwards, LDCT images having 62.5 μm pixel size are validated by the fine-tuned low-resolution model (625 μm). Obviously, network retraining can significantly improve the LDCT denoising performance, especially in reducing the alien image artifacts around the edges of the object, see

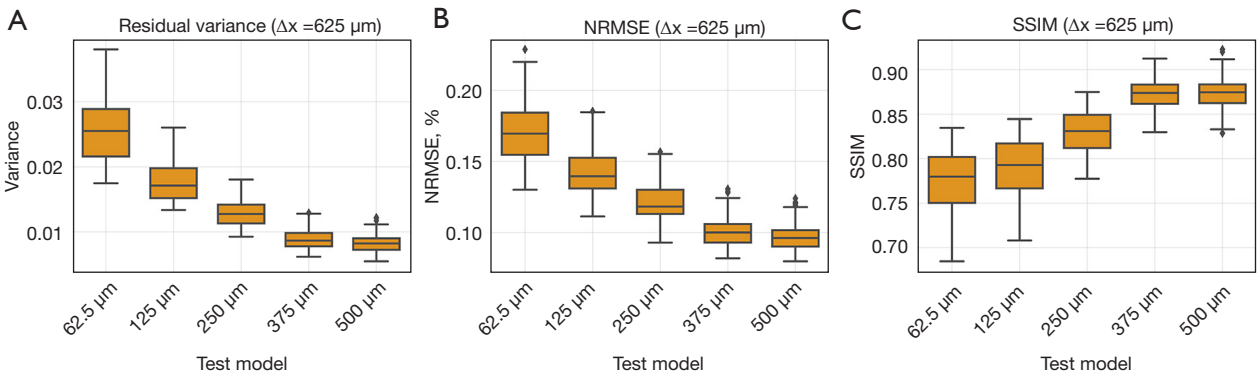

Figure 8 The quantitative results of the cross-test results of 625 μm images. (A) The variance values of residual images; (B) NRMSEs of the test images; (C) SSIMs of the test images. NRMSEs, normalized root mean square errors; SSIMs, structural similarities.

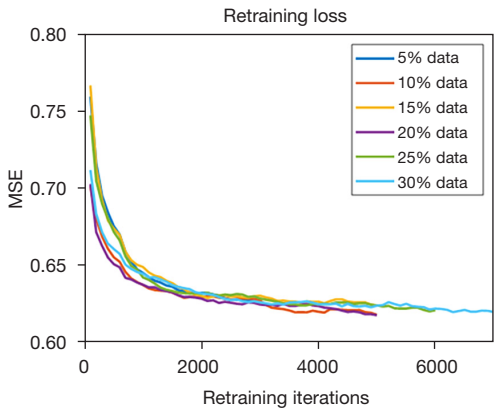

Figure 9 The U-Net retraining loss curves with six different number of retraining data. The MSE is selected as the network loss function. MSE, mean-square-error.

the zoom-in ROIs in *Figure 10*. As expected, increasing the amount of retraining data would benefit the final denoising performance, see the residual images in *Figure 10*.

Quantitative analysis results are presented in *Figure 11*. It is found that the image quality gets almost unchanged when the amount of retrained data is about 20% of the total number of training samples. Interestingly, further increasing the amount of retrained data does not result in significant improvement of image quality. Eventually, results obtained from the retrained network show close image quality to the ones obtained from the cross-tests with 125 μm model.

Noise property

In this part, noise-only CT images are studied and presented (*Figure 12*). Visually, such noise-only CT images contain unique features and look quite different at varied spatial resolutions. For instance, the noise texture on *Figure 12A* is quite uniform for CT images having 62.5 μm pixel size. Whereas, residual streaking signals gradually become apparent at the edges of the noise-only CT images as the image resolution decreases, see *Figure 12F*. We speculate that such differences in noise texture may be the root cause for blocking the transfer of U-Net between different LDCT denoising tasks with varied spatial resolutions.

Discussion

In this work, the feasibility of using the same LDCT denoising neural network, which has been trained at a certain image resolution in advance, directly onto another LDCT imaging application having different image resolution has been investigated. Herein, the most popular image-domain based neural network U-Net was employed to remove the LDCT image noise. Evaluations were performed on the data generated from the clinical CT images having six spatial resolutions.

It was found that the U-Net can efficiently remove the CT image noise for self-test and cross-test. Its denoising capability is fairly independent of the image spatial resolution. Whereas, its denoising performance is strongly correlated with the image spatial resolution (assuming the same attenuation coefficients). We guess this is mainly due to the inconsistent CT image noise textures among different image resolutions. In particular, strong streaking noise textures are easily obtained at low image resolution, regardless of the angular sampling rate. As the image resolution increases, these streaking noise textures become less significant, and more uniformed noise maps

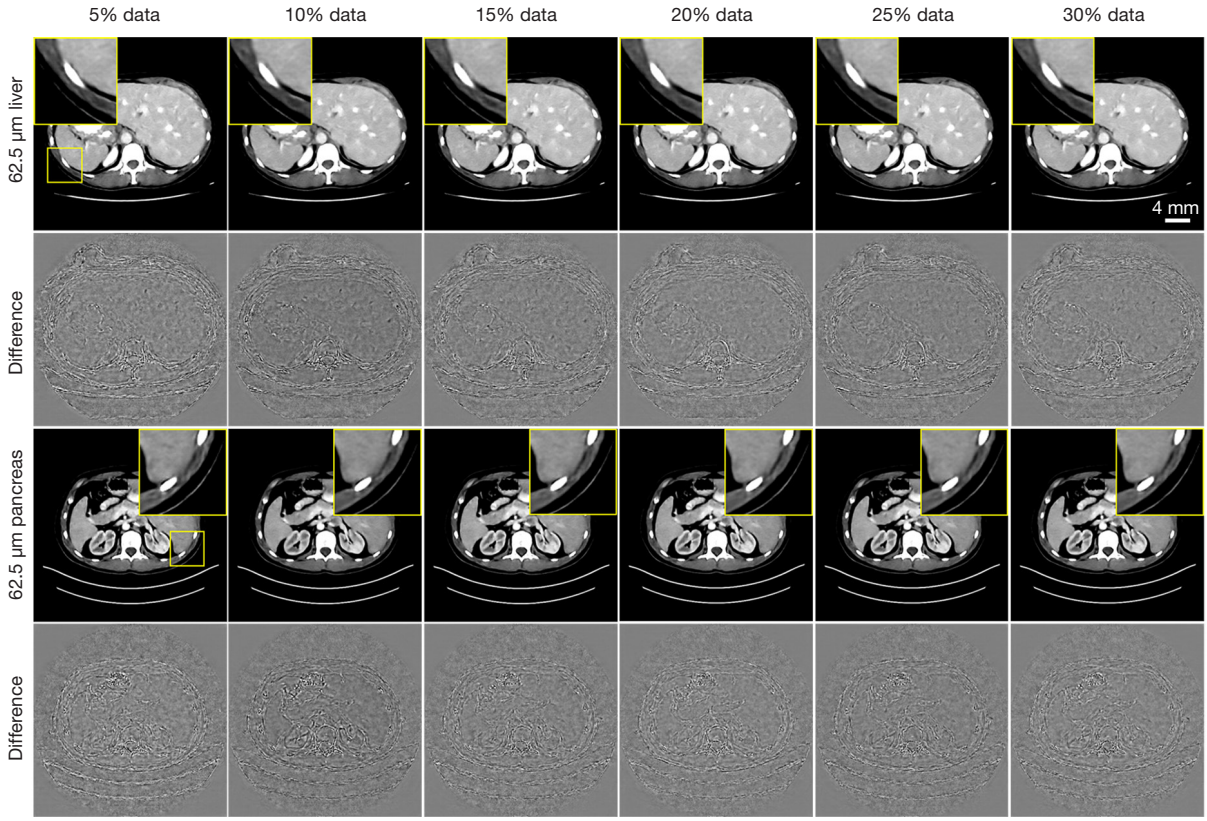

Figure 10 The U-Net retraining results of 62.5 μm images. The difference maps are obtained by subtracting the self-test images from the retraining results. The display window of test images is [0.20, 0.24] cm^{-1} and the display window of residual images is [–0.007, 0.007] cm^{-1}. The scale-bars denote 4.0 mm.

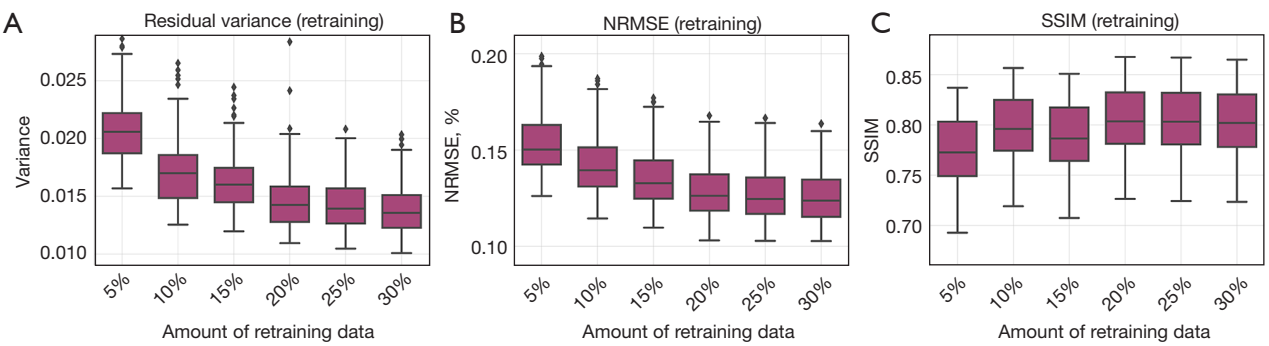

Figure 11 The quantitative results of the retraining results of 62.5 μm images. (A) The variance values of residual images; (B) NRMSEs of the retraining test images; (C) SSIMs of the retraining test images. NRMSEs, normalized root mean square errors; SSIMs, structural similarities.

are obtained. Because of these residual noise streaks, as a consequence, the immediate transformations of the U-Net between different LDCT image resolutions get challenged, especially when generating high-quality CT images. For instance, noticeable structural artifacts were presented at the edges of the object when denoising the high-resolution CT image by the U-Net trained with a lower resolution. As a contrary, in the process of denoising the low-resolution

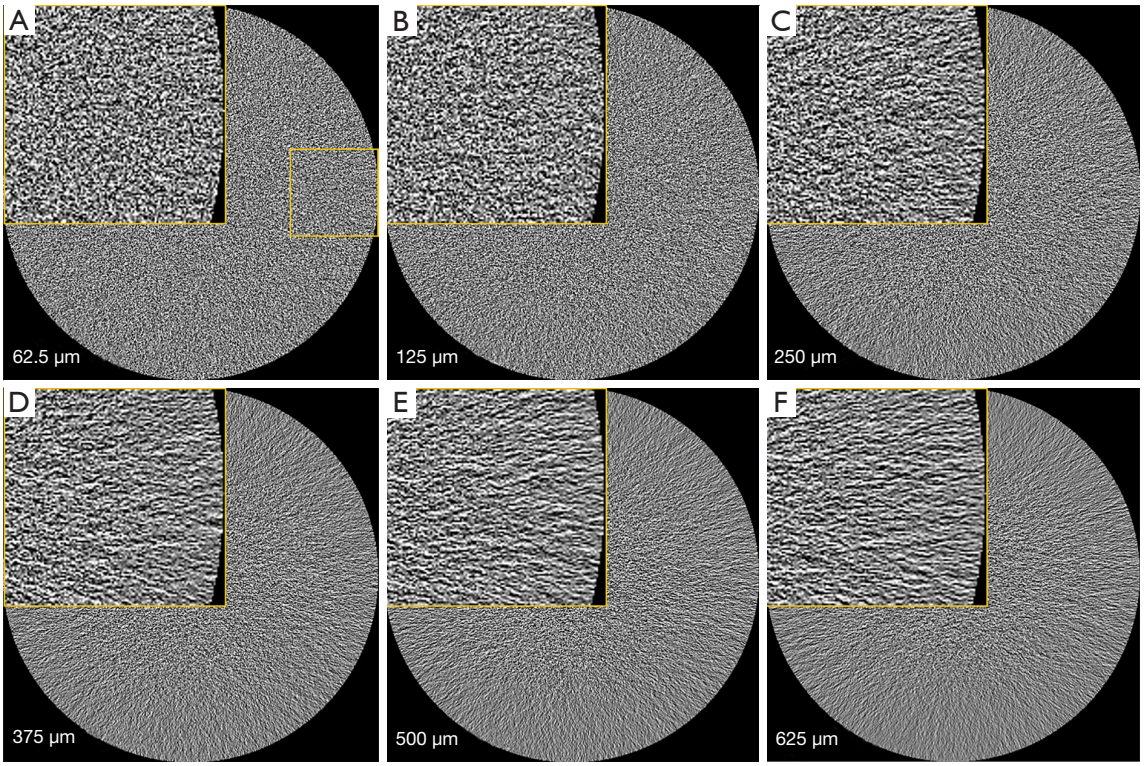

Figure 12 The noise distribution of the six different resolution CT images. The display window is [–0.026, 0.027] cm^{-1}. CT, computed tomography.

CT image by the higher resolution CT images trained U-Net, there were still some dispersion artifacts and image blurring in the middle area of the image.

Usually, the image denoising procedure would degrade the image resolution due to the blurring effect. As a consequence, balance and trade-off are often needed to remove most of the image noise while minimally impacting the image resolution. Specifically, some previous studies (22,23) have demonstrated that such image resolution degradation may depend on the object contrast: the image gets more blurred at the regions having low contrast. Despite of the important interplay between image noise and image resolution, however, this topic is beyond the main scope of this study, which only focuses on discussing the reproducibility of the denoising performance of a given network, which has already been trained in advance, among varied LDCT image datasets acquired from different imaging systems with varied resolutions. Usually, the resolution of an imaging system is defined and measured without considering the image noise.

Since it is quite difficult to completely remove such noise induced streaking features on CT images, especially for the diagnostic CT modalities (24) with spatial resolution Δx >500 μm, therefore, the current answer to the aforementioned question at the beginning of this article tends to be negative. Namely, the DL based neural network for LDCT imaging trained at one image resolution cannot be directly transferred and applied to another LDCT imaging task with a varied image resolution, even if the dose reduction ratio and the object content are similar. With the attempt to partially address this issue, network retraining was performed. It was found that only about 20% of the original samples is enough to improve the overall LDCT denoising performance. In future, if it is feasible to sufficiently remove such noise induced streaking features by utilizing novel CT image hardware, data acquisition strategies and other image reconstruction methods (25-29), we believe the same DL network could be directly transferred between different LDCT imaging applications having varied spatial resolutions without the need of additional network retraining. By that time, a great number of efforts and resources such as the data preparation and

network training can all be saved.

This study has several limitations. First, only one single DL-based CT image denoising algorithm, i.e., the U-Net, was validated. For other CT image denoising networks, e.g., GAN based DL algorithms (30-32) and 3D U-Net (33), we do not know how the results would become and careful comparisons are needed in future. Second, the different resolution data in this study were only obtained through numerical experiments of clinical CT images, and more evaluations of LDCT physical experiments need to be investigated in the future.

Conclusions

In conclusion, the DL based LDCT image denoising algorithms maybe sensitive to the CT noise textures. As a result, it is not recommended to apply the same U-Net for LDCT denoising tasks performed at different spatial resolutions to generate the same high-quality LDCT images. An optional solution is to retrain the network model with a very small amount of the target resolution data. Such transfer learning can effectively avoid introducing additional image artifacts and hence improve the model's denoising ability for LDCT denoising tasks with different resolutions.

Acknowledgments

Funding: This work was supported by the National Natural Science Foundation of China (Nos. 62201560 and 12305349), Guangdong Basic and Applied Basic Research Foundation (Nos. 2021A1515111031 and 2021TQ06Y108), Beijing Natural Science Foundation (No. 7234374), and Shenzhen Science and Technology Program (Nos. JCYJ20200109115212546 and JSGGKQTD20210831174329010).

Footnote

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://qims.amegroups.com/article/view/10.21037/qims-23-768/coif). D.L. serves as an unpaid editorial board member of *Quantitative Imaging in Medicine and Surgery*. The other authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The animal experiment was approved by the Institutional Animal Care and Use Committee (IACUC) of the Shenzhen Institute of Advanced Technology at the Chinese Academy of Sciences and was conducted in compliance with the protocol (No. SIAT-IACUC-201228-YGS-LXJ-A1498; January 5, 2021) for the care and use of animals.

References

1. Chen H, Zhang Y, Kalra MK, Lin F, Chen Y, Liao P, Zhou J, Wang G. Low-Dose CT With a Residual Encoder-Decoder Convolutional Neural Network. IEEE Trans Med Imaging 2017;36:2524-35.
2. Wu D, Kim K, El Fakhri G, Li Q. Iterative Low-Dose CT Reconstruction With Priors Trained by Artificial Neural Network. IEEE Trans Med Imaging 2017;36:2479-86.
3. Zhu B, Liu JZ, Cauley SF, Rosen BR, Rosen MS. Image reconstruction by domain-transform manifold learning. Nature 2018;555:487-92.
4. Yin X, Zhao Q, Liu J, Yang W, Yang J, Quan G, Chen Y, Shu H, Luo L, Coatrieux JL. Domain Progressive 3D Residual Convolution Network to Improve Low-Dose CT Imaging. IEEE Trans Med Imaging 2019;38:2903-13.
5. Li Y, Li K, Zhang C, Montoya J, Chen GH. Learning to Reconstruct Computed Tomography Images Directly From Sinogram Data Under A Variety of Data Acquisition Conditions. IEEE Trans Med Imaging 2019;38:2469-81.
6. Ge Y, Su T, Zhu J, Deng X, Zhang Q, Chen J, Hu Z, Zheng H, Liang D. ADAPTIVE-NET: deep computed tomography reconstruction network with analytical domain transformation knowledge. Quant Imaging Med Surg 2020;10:415-27.
7. He J, Wang Y, Ma J. Radon Inversion via Deep Learning. IEEE Trans Med Imaging 2020;39:2076-87.
8. Su T, Cui Z, Yang J, Zhang Y, Liu J, Zhu J, Gao X, Fang

S, Zheng H, Ge Y, Liang D. Generalized deep iterative reconstruction for sparse-view CT imaging. Phys Med Biol 2022.
9. Ma G, Zhao X, Zhu Y, Zhang H. Projection-to-image transform frame: a lightweight block reconstruction network for computed tomography. Phys Med Biol 2022.
10. Tamura A, Mukaida E, Ota Y, Nakamura I, Arakita K, Yoshioka K. Deep learning reconstruction allows low-dose imaging while maintaining image quality: comparison of deep learning reconstruction and hybrid iterative reconstruction in contrast-enhanced abdominal CT. Quant Imaging Med Surg 2022;12:2977-84.
11. McCollough C. TU-FG-207A-04: overview of the low dose CT grand challenge. Medical Physics 2016;43:3759-60.
12. Sidky EY, Pan X. Report on the AAPM deep-learning sparse-view CT grand challenge. Med Phys 2022;49:4935-43.
13. Touch M, Clark DP, Barber W, Badea CT. A neural network-based method for spectral distortion correction in photon counting x-ray CT. Phys Med Biol 2016;61:6132-53.
14. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer International Publishing, 2015: 234-241.
15. Pan SJ, Yang Q. A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering 2010;22:1345-59.
16. Torrey L, Shavlik J. Transfer learning. In: Handbook of research on machine learning applications and trends: algorithms, methods, and techniques. Hershey: IGI global, 2010: 242-264.
17. Weiss K, Khoshgoftaar TM, Wang DD. A survey of transfer learning. Journal of Big Data 2016;3:9.
18. Zhuang F, Qi Z, Duan K, Xi D, Zhu Y, Zhu H, Xiong H, He Q. A comprehensive survey on transfer learning. Proceedings of the IEEE 2020;109:43-76.
19. Kak AC, Slaney M. Principles of Computerized Tomographic Imaging. Philadelphia: Society for Industrial and Applied Mathematics, 1988.
20. Roth HR, Farag A, Turkbey E, Lu L, Liu J, Summers RM. Data from pancreas-CT. The cancer imaging archive. IEEE Transactions on Image Processing, 2016. doi: 10.7937/K9/TCIA.2016.tNB1kqBU.
21. Zhang X, Su T, Yang J, Zhu J, Xia D, Zheng H, Liang D, Ge Y. Performance evaluation of dual-energy CT and differential phase contrast CT in quantitative imaging applications. Med Phys 2022;49:1123-38.
22. Li K, Tang J, Chen GH. Statistical model based iterative reconstruction (MBIR) in clinical CT systems: experimental assessment of noise performance. Med Phys 2014;41:041906.
23. Li K, Garrett J, Ge Y, Chen GH. Statistical model based iterative reconstruction (MBIR) in clinical CT systems. Part II. Experimental assessment of spatial resolution performance. Med Phys 2014;41:071911.
24. Hsieh J, Nett B, Yu Z, et al. Recent advances in CT image reconstruction[J]. Current Radiology Reports, 2013, 1: 39-51.
25. Hsieh SS, Pelc NJ. The feasibility of a piecewise-linear dynamic bowtie filter. Med Phys 2013;40:031910.
26. Liu F, Yang Q, Cong W, Wang G. Dynamic bowtie filter for cone-beam/multi-slice CT. PLoS One 2014;9:e103054.
27. Willemink MJ, Persson M, Pourmorteza A, Pelc NJ, Fleischmann D. Photon-counting CT: Technical Principles and Clinical Prospects. Radiology 2018;289:293-312.
28. Flohr T, Petersilka M, Henning A, Ulzheimer S, Ferda J, Schmidt B. Photon-counting CT review. Phys Med 2020;79:126-36.
29. Leng S, Bruesewitz M, Tao S, Rajendran K, Halaweish AF, Campeau NG, Fletcher JG, McCollough CH. Photon-counting Detector CT: System Design and Clinical Applications of an Emerging Technology. Radiographics 2019;39:729-43.
30. Wolterink JM, Leiner T, Viergever MA, Isgum I. Generative Adversarial Networks for Noise Reduction in Low-Dose CT. IEEE Trans Med Imaging 2017;36:2536-45.
31. Hu Z, Jiang C, Sun F, Zhang Q, Ge Y, Yang Y, Liu X, Zheng H, Liang D. Artifact correction in low-dose dental CT imaging using Wasserstein generative adversarial networks. Med Phys 2019;46:1686-96.
32. Zhou H, Liu X, Wang H, Chen Q, Wang R, Pang ZF, Zhang Y, Hu Z. The synthesis of high-energy CT images from low-energy CT images using an improved cycle generative adversarial network. Quant Imaging Med Surg 2022;12:28-42.
33. Shan H, Zhang Y, Yang Q, Kruger U, Kalra MK, Sun L, Cong W, Wang G. 3-D Convolutional Encoder-Decoder Network for Low-Dose CT via Transfer Learning From a 2-D Trained Network. IEEE Trans Med Imaging 2018;37:1522-34.

Cite this article as: Zhang X, Su T, Zhang Y, Cui H, Tan Y, Zhu J, Xia D, Zheng H, Liang D, Ge Y. Transferring U-Net between low-dose CT denoising tasks: a validation study with varied spatial resolutions. Quant Imaging Med Surg 2024;14(1):640-652. doi: 10.21037/qims-23-768

