Atom Indonesia

Journal homepage: http://aij.batan.go.id

Noise Suppression of Computed Tomography (CT) Images Using Residual Encoder-Decoder Convolutional Neural Network (RED-CNN)

H. B. Cokrokusumo[1*], I. Hariyati[2], L. E. Lubis[1], P. Prajitno[1], D. S. Soejoko[1]

[1]*Department of Physics, Faculty of Mathematics and Natural Sciences (FMIPA), Universitas Indonesia, Depok 16424, Indonesia*
[2]*Department of Radiology, Gading Pluit Hospital, Jakarta 14250, Indonesia*

ARTICLE INFO

Article history:
Received 1 January 2021
Received in revised form 27 March 2022
Accepted 12 April 2022

Keywords:
Deep learning
Encoder-decoder network
Fully convolutional network
Image denoising
Low-dose CT
Residual network

ABSTRACT

In this study, an in-house residual encoder-decoder convolutional neural network (RED-CNN)-based algorithm was composed and trained using images of cylindrical polymethyl-methacrylate (PMMA) phantom with a diameter of 26 cm at different simulated noise levels. The model was tested on 21 × 26 cm elliptical PMMA computed tomography (CT) phantom images with simulated noise to evaluate its denoising capability using signal to noise ratio (SNR), comparative peak signal-to-noise ratio (cPSNR), structural similarity (SSIM) index, modulation transfer function frequencies (MTF 10 %) and noise power spectra (NPS) values as parameters. Evaluation of a possible decrease of image quality was also performed by testing the model using homogenous water phantom and wire phantom images acquired using different mAs values. Results show that the model was able to consistently increase SNR, cPSNR, SSIM values, and decrease the integral noise power spectra (NPS). However, the noise level on either training or testing data affects the model's final denoising performance. The lower noise level on testing data images tends to result in over-smoothed images, as indicated by the shift of the NPS curves. In contrast, higher simulated noise level tends to result in less satisfactory denoising performance, as indicated by lower SNR, cPSNR, and SSIM values. Meanwhile, the higher noise level on training data images tends to produce denoised images with reduced sharpness, as indicated by the decrease of the MTF 10 % values. Further studies are required to better understand the character of RED-CNN for CT noise suppression regarding the optimum parameters for best results.

INTRODUCTION

Clinical imaging practices using computed tomography (CT) must be performed by complying with the 'as low as reasonably achievable' (ALARA) principle. This is because excessive exposure to X-ray radiation would increase the probability of cell damage, which should be avoided [1,2]. On the other hand, too low X-ray exposure would result in low image quality, which is also undesirable as noise would become more prominent when the image receptor receives low signal. The relation between exposure and noise has made low noise images to be associated with the use of high exposure. While this resounds the ALARA and optimization principle, efforts can still be made by reducing the noise on the images acquired using low exposure as part of optimization process.

One possible method for this is to employ artificial intelligence (AI) methods, including the convolutional neural network (CNN) to denoise CT images. AI-based methods, like all image processing methods, promise less need for X-ray exposure as they only deal with image data. One example used in this study is the residual encoder-decoder convolutional neural network (RED-CNN) proposed by Chen et al. [3]. In their study, RED-CNN's denoising results suggest better

*Corresponding author.
E-mail address: haryo.bimo@sci.ui.ac.id
DOI: https://doi.org/10.17146/aij.2022.1113

denoising performance compared to other image denoising methods. The results' evaluation is based on the similarity of RED-CNN-processed noisy low-dose images to original normal-dose images. These evaluation parameters include comparative peak signal-to-noise ratio (cPSNR) and structural similarity (SSIM) index [4,5]. RED-CNN is shown to give results closer to less noisy original images than other algorithms mentioned.

In this study, additional in-depth analysis of sharpness and noise texture was also done. While this was discussed by Chen et al. [3], their evaluation was limited to qualitative measures. The modulation transfer function (MTF) cut-off frequencies and noise power spectra (NPS) were used to evaluate these factors quantitatively [6]. MTF frequencies were used to evaluate image sharpness while NPS curves were used to evaluate the resulting image's noise properties. These evaluations were needed as the resulting images may suffer from over-smoothing since some CNN architectures do not capture image texture well [7]. Signal-to-noise ratio (SNR) [8], which gives a rough idea of the smoothness of an image, was also considered.

METHODOLOGY

In-house phantom images were used as both training and testing dataset. For training dataset, 26 cm PMMA cylindrical phantom images, obtained using Philips Ingenuity CT (Philips Healthcare, Best, the Netherlands), were used. These images have a dimension of 400 × 400 mm represented as 512 × 512 pixels images. These images were obtained at mAs value of 210 mAs and a slice thickness of 0.625 mm at both 120 kVp and 140 kVp peak tube voltage. Another geometrical variation of this phantom (21 × 26 cm elliptical PMMA phantom) was also used as an additional testing dataset for verification.

Besides the PMMA phantom images, other phantoms were also used for verification as testing dataset. These phantoms include homogenous water phantom for resulting noise texture evaluation and wire phantom for resulting sharpness evaluation using MTF measurements. Both phantoms have 32 cm diameter and were measured at different mAs values (155 mAs, 200 mAs, 250 mAs, and 300 mAs) to bring variation in the noise level of images used as testing dataset. The peak tube voltage of these images was fixed at 120 kVp with a slice thickness of 8 mm. The images were obtained using Siemens SOMATOM Cardiac 64 (Siemens Healthineers, Erlangen, Germany).

Due to lack of dose level variation in PMMA phantom dataset, noise simulation for these images is required. The simulation can be mathematically written as Eqs. (1) and (2):

$$I_{ld,sim} = \text{Poisson}\left(\bar{I}_{ld,sim}\right) + \text{Gaussian}(m_e, \sigma_e) \quad (1)$$

Where

$$\bar{I}_{ld,sim} = I^0_{ld,sim}\exp(-p_{nd}). \quad (2)$$

p_{nd} represents sinogram data of the images obtained using fan-beam projection data and attenuation coefficients calculation from HU values in the images [9]. $I^0_{ld,sim}$ represents blank scan flux, which can be changed to represent different noise levels on images [10,11]. Greater $I^0_{ld,sim}$ value would represent higher photon flux which indicates a lower noise level [3]. For a variation in noise level used, $I^0_{ld,sim}$ values of 5.0×10^4, 7.5×10^4, 1.0×10^5, 1.5×10^5, and 2.0×10^5 were chosen for the training dataset, while values of 5.0×10^4, 1.0×10^5, 1.5×10^5, 2.5×10^5, and 5.0×10^5 were chosen for the testing dataset. All testing dataset variation measurements were performed with a model trained using PMMA phantom images with a blank scan flux value of 1.5×10^5, while all training dataset variation measurements were performed on PMMA phantom images with a blank scan flux value of 5.0×10^4. m_e and σ_e represent mean and variance of electronic noise which are discarded in this study as invasive measurement on CT detector system is required. Once noise was added to the images, the resulting sinogram data would then be reconstructed using fan-beam reconstruction method with the Ram-Lak filter [3,7]. The noisy images are then used as RED-CNN's input, with original images being the ground truth.

Data augmentation was required to enhance the generalization capability of the trained model. Transformation operations, such as 90˚ rotation and mirroring, were applied. The resulting transformations would result in eight times the amounts of original data. Due to limited availability of images available (only four images, two images at 120 kVp and the other two at 140 kVp), the usual transformation operations were not sufficient, and an additional augmentation method, which was patch extraction (Fig. 1), was required. This method works

by dividing images into equally sized smaller parts to increases the dataset's size, allowing the network to learn local details required for optimal noise suppression [3]. In this study, a patch size of 64 × 64 pixels was used, with a stride of 4 pixels and no padding. This setting was applied to the cropped image dataset (cropped from 512 × 512 pixels to 384 × 384 pixels). It dramatically increased the dataset's size to over 204,800 unique noisy-normal patch pairs extracted from only four original images. 80 % of noisy-normal patch pairs were used as the training set while the other 20 % as the validation set. These 204,800 patch pairs were then trained using the RED-CNN model.

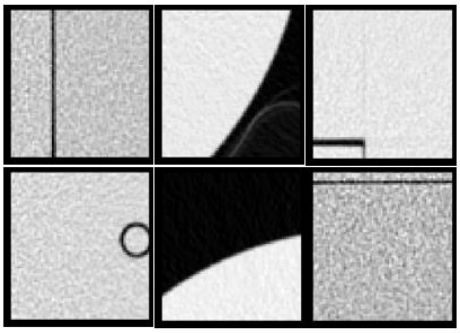


Fig. 1. Examples of extracted patches to be used as the training dataset.

The configuration of the CNN architecture used in this study is shown in Fig. 2 [3]. The architecture consists of five convolution layers and five transposed convolution layers with residual connections connecting the two types of layers. The convolution layers work as the encoder part of the network, while the transposed convolution layers work as the decoder part. The convolutional and transposed convolutional layers have 96 filters with a size of 5 × 5, a stride of size 1, and zero padding. This configuration does not include the first convolutional layer and the last transposed convolutional layer, both of which only have one filter. Convolutional layers would reduce noise and artifact in a CT image. However, their use would be followed by some structural details loss that can be compensated by using transposed convolutional layers that work to recover structural details of the images used [3]. Rectified linear unit (ReLU) function is used in both the decoder and the encoder part of the network, so the sparsity of the features can be maintained [12]. Lastly, residual connections are implemented to enhance the loss optimization process as accumulated loss in the network may result in unsatisfactory denoised images [3].

The RED-CNN model was trained using Adam loss optimizer with optimizer parameters β_1 and β_2 set to β_1 = 0.9 and β_2 = 0.999 (Kingma & Ba, 2015). The initial learning rate was set to 0.0001, which slowly decayed by a rate of 0.991 every four epochs. This training process was performed for 40 epochs. The loss function used for the optimization process was mean squared error loss function which can be mathematically written as Eq. (3):

$$\text{Loss} = \frac{1}{N} \sum_i^N \left|\left| Y_i - M(X_i) \right|\right|^2 \quad (3)$$

where Y_i and X_i are normal dose and low dose CT image patches respectively and M is the mapping function applied to low dose inputs to produce outputs as close as possible to the real normal dose images [3]. This training process used patch data as both the input and the ground truth. As a result, image patch reconstruction needed to be performed. The reconstruction of the images was done by taking the average values of the overlapping area of the patches [13].

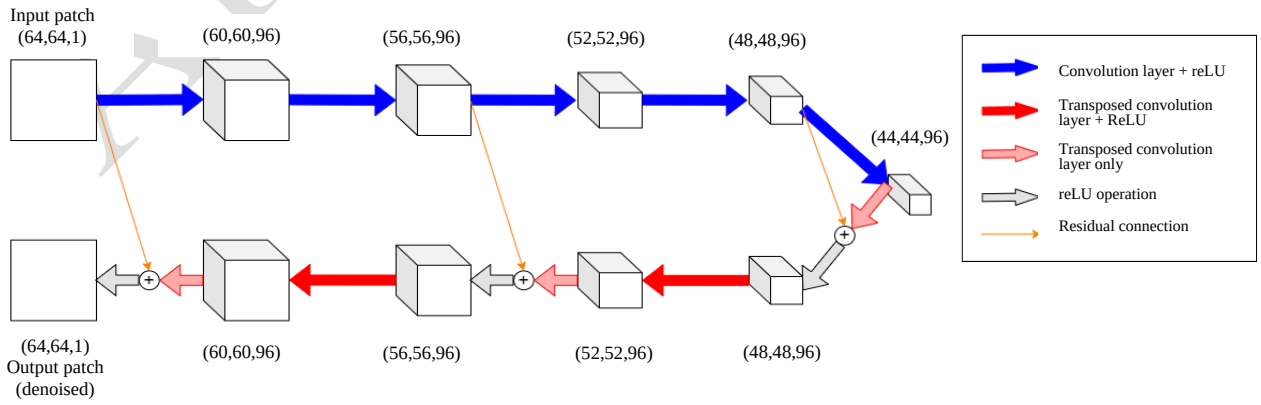


Fig 2. RED-CNN architecture used for CT image noise suppression [3].

Five methods of evaluation were used in this study. These include SNR, cPSNR, SSIM, MTF 10 %, and NPS. SNR measurements were done by selecting five different regions of interest (ROI) on the phantom, which should exclude area with a pictured bead. One example of SNR measurements is shown in Fig. 3 The same ROI selections can also be used to measure the NPS curves. Meanwhile, cPSNR and SSIM measurements were done using all areas of the phantom depicted in the images, although the images used for these measurements were cropped to the size of 384 × 384 pixels to exclude most blank area in the images. cPSNR and SSIM measurements were done only on PMMA phantom images as these measurements need a reference of how a denoised image should look like and only PMMA phantom images whose noise level is modified. Lastly, MTF spatial frequencies measurements were done only on ROI with a bead depicted as the bead is the only object in the phantom images to represent the point spread function (PSF) of an imaging system [14]. All measurements also considered cross-validation to give a clearer idea of the consistency in the measurement results when the training dataset's composition is changed [15].

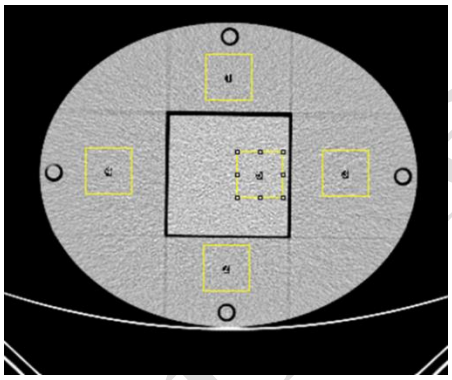


Fig. 3. ROIs used for SNR and NPS measurements.

RESULTS AND DISCUSSION

The denoising results for different testing noise levels (PMMA phantom samples) are shown in Fig. 4. The results show that all the image samples denoised using RED-CNN have their noise level reduced. However, the noise improvement is still affected by the noise level in original image samples. It is seen that the resulting images from the noisiest samples result in improved images with higher noise levels compared to other resulting images from less noisy original images. This is also supported by the measurement results of some evaluation parameters such as SNR and NPS.

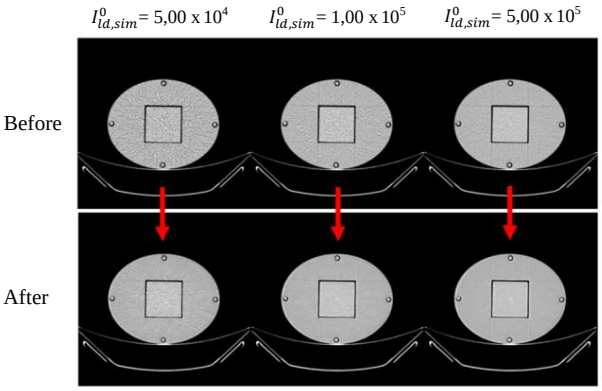
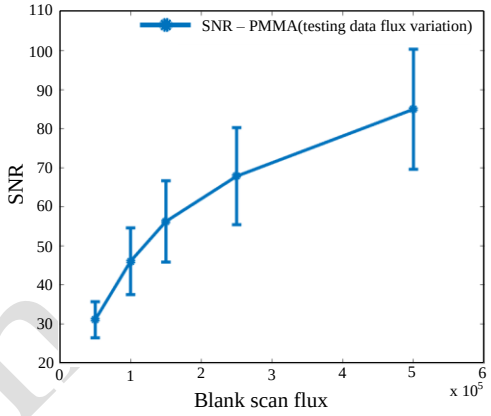
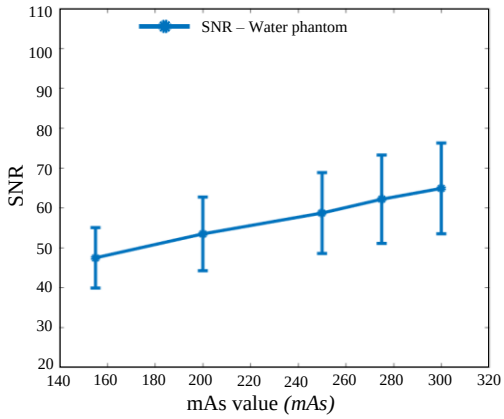


Fig. 4. The denoising results of the images on different testing noise levels.



(a)



(b)

Fig. 5. SNR measurements of denoised (a) PMMA phantom images (test dataset) and (b) water phantom images.

SNR measurements show a consistent increase in SNR values when the blank scan flux value of the testing dataset is increased, as shown in Fig. 5, which means consistent decreases in the input images' noise level. These measurements were tested using PMMA and water phantom images. The noise level uses actual mAs values used during the data acquisition process for water phantom images. For PMMA images, the noise level is indicated by blank scan flux value used for synthetic noise simulation.

These measurements mean the model can denoise entirely new phantom images (i.e., water phantom images and different geometries of phantoms with the same materials). Despite that, there is a difference in how the SNR value increases for other materials that might result from exclusively using the PMMA dataset in the training process.

Additionally, selecting the optimal blank scan flux in the training dataset is needed. Results show that a decrease of the SNR value is possible in flux values that are either too high or too low as shown in Fig. 6. In this study, the maximum SNR value measured for noise level variation in the training dataset is at a blank scan value of 7.5×10^4. This value may change as the training dataset used is also changed.

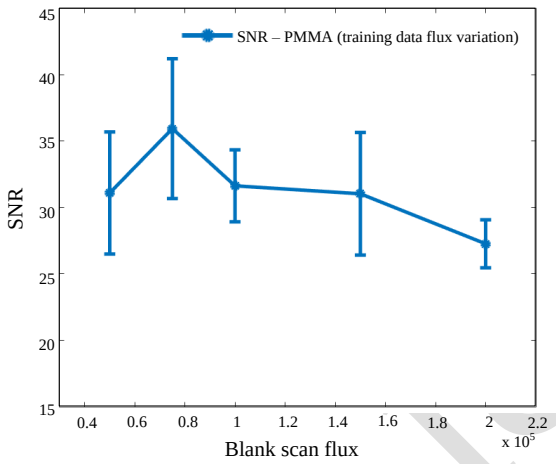


Fig. 6. SNR measurements of denoised PMMA phantom images with different noise level in the training dataset.

MTF frequency measurements show no significant change to images' sharpness when the RED-CNN algorithm was applied to the test dataset (PMMA and wire phantom images) with varying noise levels indicated by synthetic blank scan flux values for PMMA images and real mAs values for wire phantom images (Fig. 7), similar to noise levels indicators in SNR measurements. Significant reduction in MTF frequency only happens with the use of highly noisy training dataset, as shown in Fig. 8. An up to 45.41 % reduction of MTF 10 % frequency value is measured at the lowest flux value used (5.0×10^4). Therefore, highly noisy images are not suitable to be used as the training dataset as the resulting denoised images depend significantly on the training dataset quality.

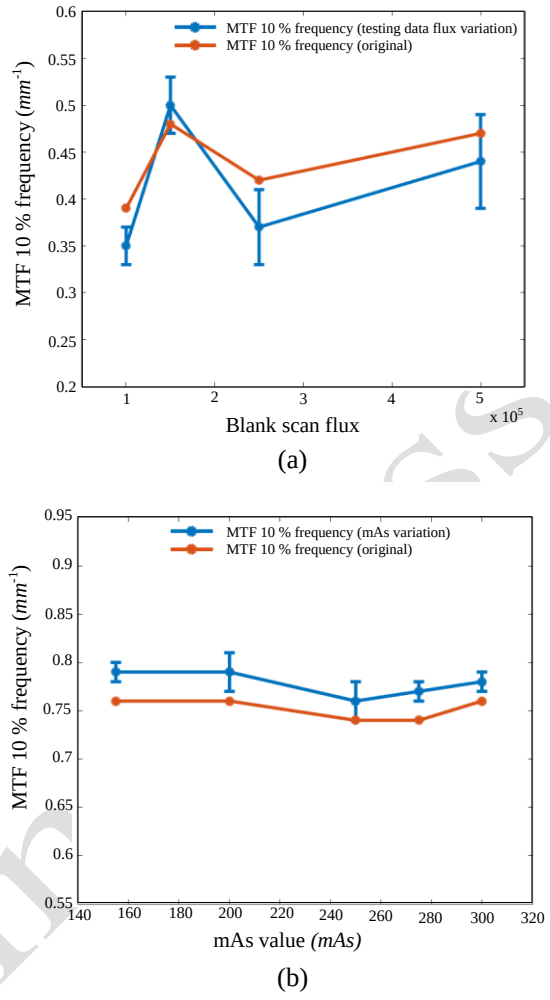


Fig. 7. MTF 10 % measurements with different testing noise levels using (a) PMMA phantom dataset and (b) wire phantom dataset.

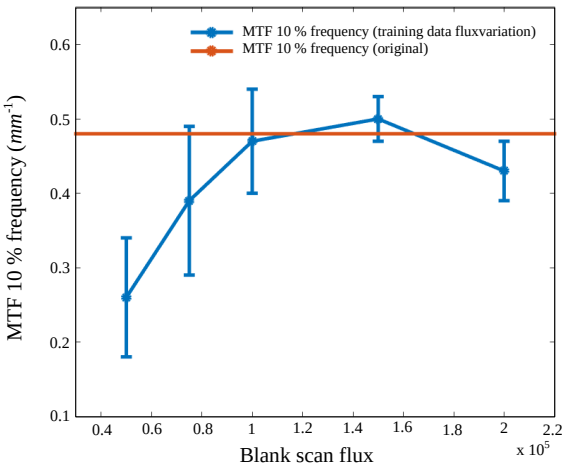


Fig. 8. MTF 10 % measurements done for different training data noise levels.

NPS measurements for the denoising results are shown in Fig. 9 and Fig. 10. The NPS curves show consistent decreases in denoised images' noise levels. Like SNR measurements, lower testing images' noise would result in denoised images with

a lower noise level, as marked by the NPS curves' flattening. However, the resulting NPS curves have resulted in a new NPS peak at lower frequency values. This is more prominent in image data with lower noise levels. This means that as the test images' noise level decreases, the noise texture in the resulting denoised images changes and the resulting images become more prone to over-smoothing. The change of the NPS peak to lower frequency values would indicate image over-smoothing. Thus, RED-CNN is not suitable for use on CT images that are already less noisy.

cPSNR and SSIM measurements for noise level variations in both training and testing datasets are shown in Fig. 11 and Fig. 12. These measurements show an increasing similarity of the denoised images to the low-dose reference with no noise added as the blank scan flux value was increased in the testing dataset. However, both cPSNR and SSIM measurements performed show a consistent decrease in the difference between noisy and denoised values. This could indicate the model's decaying effectiveness to produce denoised images as the flux is increased. At the highest blank scan flux value used (5.0×10^5), the resulting SSIM value does not differ much from the noisy images', meaning that there is almost no improvement in the resulting images' similarity to the reference low-dose images. This may be caused by changes in noise texture, as indicated by the NPS curves' changes. Measurements of cPSNR and SSIM on training data variation also show a tendency of saturation in the values measured on mid-level blank scan flux, which is parallel to the SNR value trend on different training data noise level.

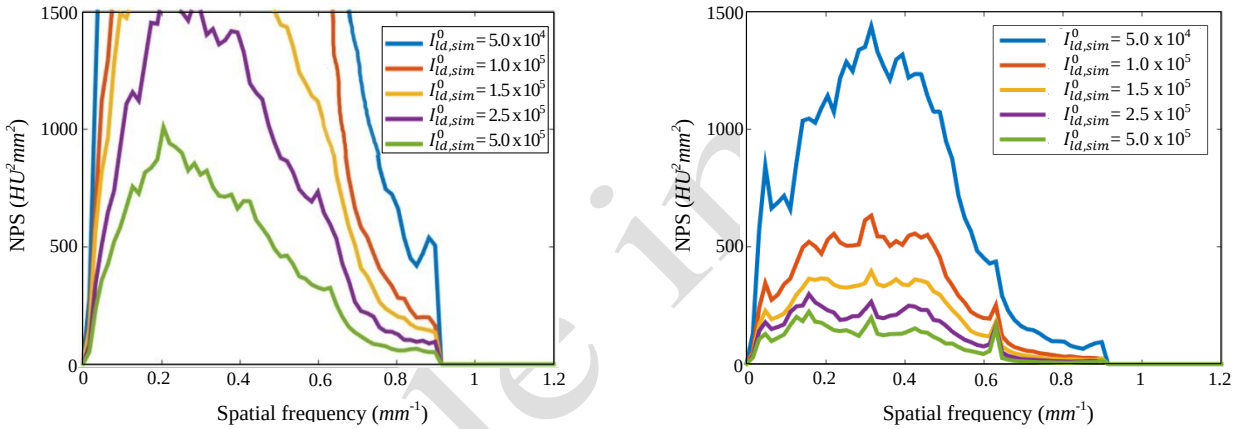


Fig. 9. NPS curves of PMMA phantom images (a) before and (b) after denoising.

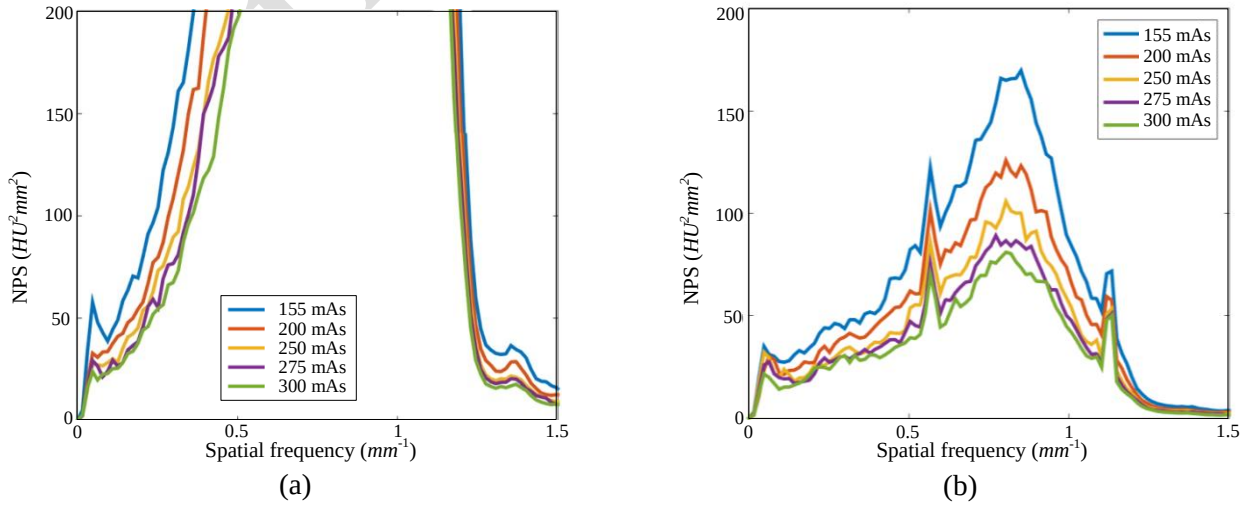


Fig. 10. NPS curves of water phantom images (a) before and (b) after denoising.

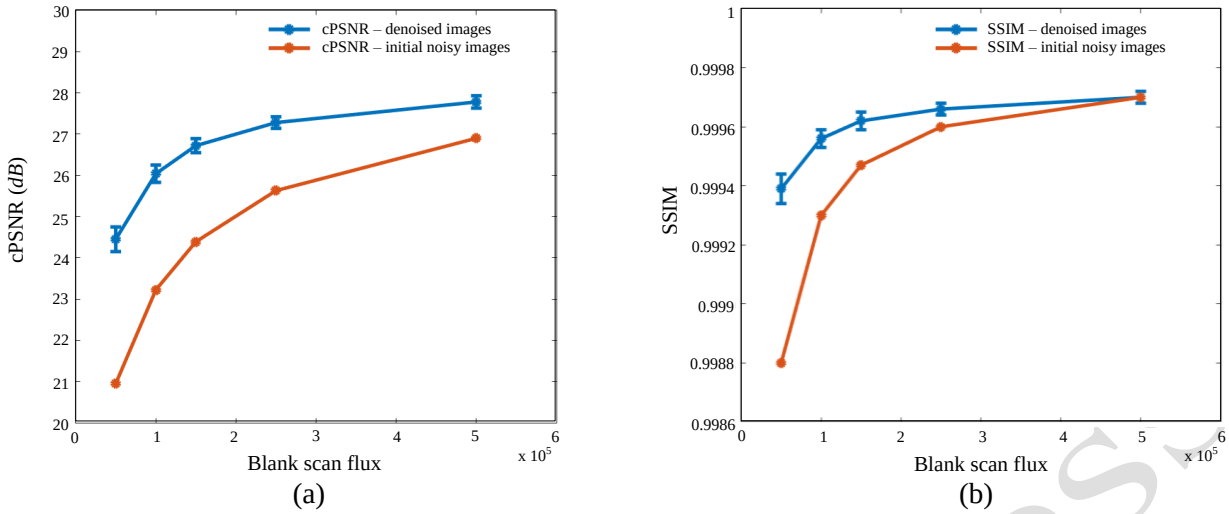


Fig. 11. cPSNR (a) and PSNR (b) measurements for variation in the testing data noise level.

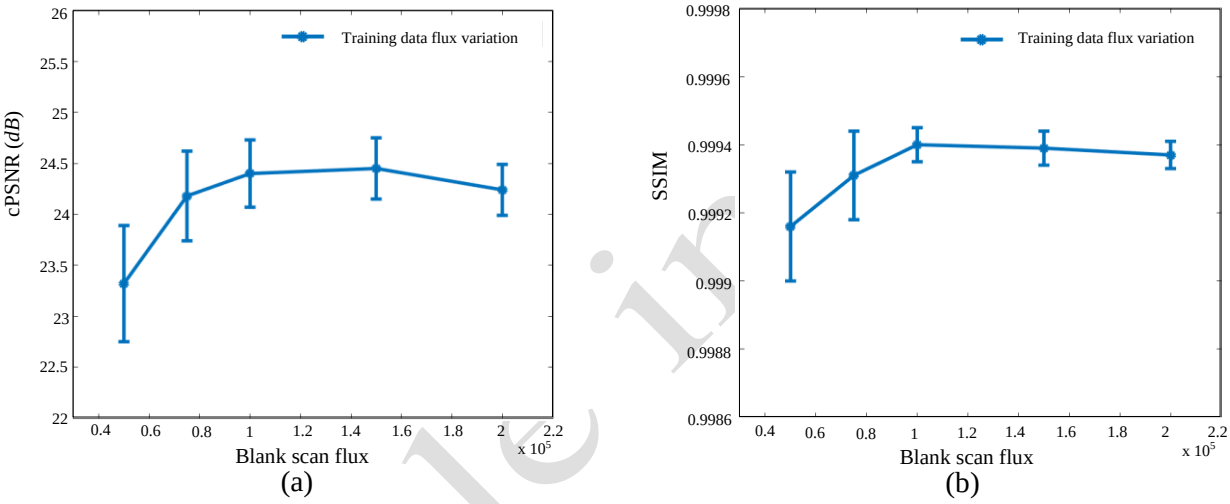


Fig. 12. cPSNR (a) and PSNR (b) measurements for variation in the training data noise level.

In general, the study shows that although the use of RED-CNN had presented with limitations regarding the noise property of the training and testing images, it demonstrated a major potential towards low-dose CT imaging. The study has not yet considered the usage of RED-CNN on actual clinical data as it demands further ethical clearance. Subsequent studies using real clinical data are strongly suggested to better present the possibility of using this method in a real clinical scenario.

CONCLUSION

It can be concluded that the denoising performance of the model depends on the noise level of both the training and testing dataset used. As the noise level on the testing dataset is reduced, the model's performance becomes more prone to over-smoothing, and the enhancement of the images' similarity to a low dose reference becomes saturated. By varying the noise level on the training dataset, it is also shown that applying too high training noise level would result in decreased image sharpness while using too low training noise level would result in reduced similarity (to normal-dose images) and SNR in resulting images. Further study to investigate the optimum set parameters for best results are required.

ACKNOWLEDGMENT

The author would like to thank Indah Lestariningsih from Cibinong Regional General Hospital who contributed to the availability of a part of image data used in this study.

AUTHOR CONTRIBUTION

H. B. Cokrokusumo contributed as the main contributor of this paper. All authors read and approved the final version of the paper.

REFERENCES

1. J. T. Bushberg, J. A. Seibert, E. M. Leidholdt *et al.*, The Essential Physics of Medical Imaging, 3rd ed., Lippincott Williams & Wilkins, Philadelphia (2012) 852.
2. D. R. Dance, S. Christofides, A. D. A. Maidment *et al.*, Diagnostic Radiology Physics: A Handbook for Teachers and Students, International Atomic Energy Agency, Vienna (2014) 500.
3. H. Chen, Y. Zhang, W. Zhang *et al.*, Biomed. Opt. Express **8** (2017) 679.
4. A. Horé and D. Ziou, Int. Conf. Pattern Recognit. (2010) 2366.
5. Z. Wang, A. C. Bovik, H. R. Sheikh *et al.*, IEEE Trans. Image Process. **13** (2004) 600.
6. S. N. Friedman, G. S. K. Fung, J. H. Siewerdsen *et al.*, Med. Phys. **40** (2013) 1.
7. B. Kim, M. Han, H. Shim *et al.*, Med. Phys. **46** (2019) 3906.
8. A. R. Goode, C. Snyder, A. Snyder *et al.*, J. Appl. Clin. Med. Phys. **20** (2019) 172.
9. J. M. Boone and J. A. Seibert, Med. Phys. **24** (1997) 1661.
10. J. Ma, Z. Liang, Y. Fan *et al.*, Med. Phys. **39** (2012) 4051.
11. D. Zeng, J. Huang, Z. Bian *et al.*, IEEE Trans. Nucl. Sci. **62** (2015) 2226.
12. X. Glorot, A. Bordes and Y. Bengio, Proc. 14th Int. Conf. Artif. Intell. Stat. **15** (2011) 315.
13. C. Wang, Z. Zhao, Q. Ren *et al.*, Entropy **21** (2019) 1.
14. K. Rossmann, Radiol. **93** (1969) 257.
15. R. Kohavi, Int. Joint Conf. Artif. Intell. **14** (1995) 1137.