Original Article

Neural Network Performance Evaluation of Simulated and Genuine Head-and-Neck Computed Tomography Images to Reduce Metal Artifacts

Abstract

Background: This study evaluated the performances of neural networks in terms of denoizing metal artifacts in computed tomography (CT) images to improve diagnosis based on the CT images of patients. **Methods:** First, head-and-neck phantoms were simulated (with and without dental implants), and CT images of the phantoms were captured. Six types of neural networks were evaluated for their abilities to reduce the number of metal artifacts. In addition, 40 CT patients' images with head-and-neck cancer (with and without teeth artifacts) were captured, and mouth slides were segmented. Finally, simulated noisy and noise-free patient images were generated to provide more input numbers (for training and validating the generative adversarial neural network [GAN]). **Results:** Results showed that the proposed GAN network was successful in denoizing artifacts caused by dental implants, whereas more than 84% improvement was achieved for images with two dental implants after metal artifact reduction (MAR) in patient images. **Conclusion:** The quality of images was affected by the positions and numbers of dental implants. The image quality metrics of all GANs were improved following MAR comparison with other networks.

Keywords: *Denoizing, head-and-neck cancer, metal artifacts, neural networks*

**Goli Khaleghi¹,
Mohammad Hosntalab¹,
Mahdi Sadeghi²,
Reza Reiazi²,³,
Seied Rabi Mahdavi²**

¹*Department of Medical Radiation Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran, ²Department of Medical Physics, School of Medicine, Iran University of Medical Sciences, Tehran, Iran, ³Princess Margaret Cancer Center, Toronto, Ontario, Canada*

Introduction

Complicated mechanisms such as beam-hardening effects and photon starvation lead to metal artifacts in computed tomography (CT) images.[1] High-density materials produce dark bands and/or streaking artifacts and severely reduce the quality of reconstructed images.[2,3] In particular, dental fillings lead to incorrect estimations of anatomical structures and CT numbers, leading to imprecise dose calculations for head-and-neck radiotherapy.[4] In treatment planning, metal artifact regions are manually defined and may be replaced by water to minimize their effects based on density correction.[5] However, this task is difficult and laborious, and a substantial interobserver variability may occur in manual tumor delineation, leading to errors in the calculations of doses.[6,7] Deep learning is a new method to reduce the number of metal artifacts in CT images. Deep learning has begun to be used extensively in recent years to handle many complicated tasks.[8]

Researchers attempted a straightforward method to use density overrides in the Pinnacle Treatment Planning Software (Philips Healthcare). In the planning system called Monaco (ELEKTA), a Monte Carlo algorithm is used to identify artifacts derived from the use of high-density materials. The developers of Monaco contrasted the dosage calculation algorithms of two systems on CT images of patients without adjusting for either implant densities or encompassing tissues.[9] Convolutional neural networks (CNNs) are organized by metal artifact reduction (MAR) systems, where data of initial and corrected images are fused to remove artifacts.[10] In the damage stages, precorrected and uncorrected images were applied as input for the prepared CNN to create CNN images with reduced artifacts. The results demonstrated that deep learning could serve as a new means of addressing the reconstruction challenges of CT and may lead to a more precise estimation of tumor volumes for planning radiation treatment.[11] In another study,[12] a method

Address for correspondence:
Prof. Mahdi Sadeghi, Department of Medical Physics, School of Medicine, Iran University of Medical Sciences, P.O. Box: 14155-6183, Tehran, Iran.
E-mail: sadeghi.m@iums.ac.ir

How to cite this article: Khaleghi G, Hosntalab M, Sadeghi M, Reiazi R, Mahdavi SR. Neural network performance evaluation of simulated and genuine head-and-neck computed tomography images to reduce metal artifacts. J Med Sign Sens 2022;12:269-77.

derived from a conditional generative adversarial neural network (cGAN) for reducing the metal artifacts in CT ear images of cochlear implant recipients was developed. Researchers tested cGAN using post-implantation CT images of 74 ears and quantitatively assessed the quality of artifact-corrected images by contrasting the segmentations of intracochlear anatomical structures in the genuine pre-implantation with artifact-corrected CT images. In a final study,[13] a useful strategy that employs GAN was proposed to reduce noise and protect texture elements in images. Several rehashed densely sampled B-scan optical coherence tomography (OCT) images were utilized with multi-frame enlistment to produce a denoizing generator.

In the current study, head-and-neck phantom images were simulated using MATLAB software. The artifacts were generated by dental implants (in random locations on the phantom images). After all, datasets were used as inputs in six types of neural networks using quality image metrics as well as loss and accuracy plots which is the novel approach presented in this study. The noisy and noise-free head-and-neck CT images of patients were simulated by MATLAB to increase the amount of data, and the images were imported to train and validate the GAN network. MAR by the GAN network was affected by the positions and numbers of dental implants, which was another novelty. The limited quantity of input data for training and validating neural networks represents a limitation of this study.

Materials and Methods

Studies on simulated phantom images and patients first were described. Six types of neural networks were compared based on quality image metrics to choose the most efficient network for MAR in the patient study

Simulated phantom images

Head-and-neck phantom images were simulated by MATLAB R2019a (Math Works, MA, USA). The row and column numbers in the phantom images were specified as positive integers. Six numerical matrices were used to characterize ellipse parameters for the phantom images. For any given pixel within the yield image, the pixel's value was increased to the entirety of the added substance intensity values for all circles of which the pixel could be a part, and when a pixel was not a part of any ellipse, its value was zero.[14] When metal implant and teeth densities were considered, dental implant densities on the head-and-neck phantoms were equal to 1.3 times the number of teeth in the phantom matrix in MATLAB.[15] Images that randomly showed an implant in the pixelated area of the dental segmentation were created. With an implant considered part of the dental segmentation in random positions, 3000 images were created as neural network training data for the first step. For the next step, the radon function in MATLAB was used to transform those images into images with

artifacts. Then, 600 noise-free images and 600 images with dental artifacts were randomly created to validate the neural networks. With three dental implants considered as part of the dental segmentation with random positions, 200 images were created. Figure 1 shows head-and-neck phantom images with one implant for training and validation, three implants for testing, and one without any implant.

Neural network modeling

The data were transmitted to Google Colaboratory notebook by importing Keras, Numpy, Skimage, Matplotlib. Pyplot, Glob, and Tensorflow libraries. Codes for an autoencoder (AE), generative UNet, denoizing CNN (DnCNN), residual network (ResNet), visual geometry group (VGG) (a group of researchers at Oxford who developed this architecture) VGG, and GAN were separately trained by 3000 noisy images and 3000 noise-free images. These codes were validated by 600 noise-free images and 600 noisy images. Finally, networks were tested by 200 noise-free and 200 noisy images.

All networks were designed to correct images with low artifacts as output and down-sampled to a final 256 × 256 resolution. With rectangular images considered, the images were rescaled, and the central patch of 256 × 256 was cut from the subsequent images for all networks. The losses and accuracies of the networks were compared in separate diagrams [Figure S1]. Networks were trained and run several times to identify the best hyperparameters for each network according to loss, accuracy, and image quality indices comparison. Hyperparameters were changed in each network and the architectures were developed to be appropriate for the MAR process which has mentioned in the introduction of each network. We developed networks and changed structures to make them unique for our study. Comparing these networks using image quality measurements is the validation of choosing the best network for patient study while has not done in other researches.

Denoizing convolutional neural network architecture

The DnCNN computes the difference between noisy and latent clear images. The CNN in this study had 20 layers, including rectified linear unit (ReLU) activation, batch
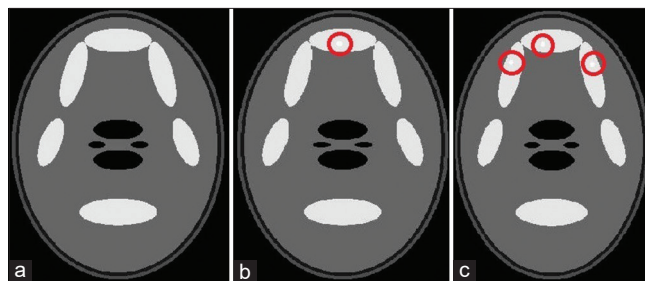

Figure 1: Images simulated by MATLAB; (a) Original phantom image with no dental implant; (b) Phantom image with one dental implant, and (c) Phantom image with three dental implants

normalization, and regression output layers with 1×1 strides and 64 kernel filters of 3×3 in size. The loss function is defined as Eq. 1:[16]

Loss = (denoized images– noise-free images)^{½}/2 (1)

Auto encoder architecture

To build an AE, three components are required: Encoding and decoding methods as well as a loss function to compare the output and target. In this network, an encoder layer with eight latent dimensions, three kernel filters of 16×32 in size, two strides, and exponential linear unit activation and convolutional two dimensional (2D) layer were used. The decoder layer was the same as the encoder used when the convolution layer was transposed. Latent dim was equal to 8, which shows the number of layers in the network. The mean square error was used as a loss function, which was imported from the Keras library.[17,18]

UNet architecture

The UNet architecture consists of two major sections and is symmetric. The first section is known as the contracting path and is based on the general convolutional process. The second part is the expansive path, which is developed using transposed 2D convolutional layers. Five layers are used in the first part, including convolutional 2D, batch normalization, ReLU activation, and drop-out layers. It has maxpooling2D of 2×2 in size, and 16, 32, 64, 128, 256 kernel filters of 3×3 in size. Similarly, five layers are used in the second part. However, in the second part, convolutional 2D transpose is used with the concatenating layers using similar information as in the first part. The drop-out factor of each layer was 0.1, 0.1, 0.2, 0.2, and 0.3, respectively, which helps the network to delete extra information. The loss function as a binary cross entropy was imported from the Keras library.[19,20]

VGG16 architecture

VGG16 is a CNN model. Our network consisted of five blocks with the number of 16, 32, 64, 128, 512 kernel filters of 3×3 in size, convolutional 2D fully connected layers with drop-out and maxpooling2D of 2×2 in size, and bilinear upsampling 2D of 32×32 in size. The drop-out factor of each block was 0.1, 0.1, 0.2, 0.2, and 0.5, respectively. ReLU activation layers were used, and the stride was 2×2. The mean square error was used as a loss function imported from the Keras library.[21,22]

ResNet architecture

This network consisted of five blocks and each block consisted of bit channels, convolutional layers with a drop-out layer, four kernel filters with (4, 4, 16), (8, 8, 32), (16, 16, 64), (32,32, 128) in size and bilinear upsampling 2D of 32×32 in size. The drop-out factor of each block was 0.1, 0.1, 0.2, 0.2, and 0.2, respectively. The mean square error was utilized as a loss function imported from the Keras library.[23,24]

Generative adversarial neural networks architecture

As an extraordinary model of the neural network, the GAN includes two networks that are trained simultaneously: One focused on image generation, the other centered on segregation. The Pix2Pix algorithm was previously proposed for performing image-to-image translation.[25,26] The generative network continues generating images that approximate the genuine images when the discriminator attempts to recognize the contrasts between fake and genuine images. It can then denoize the artifacts. The GAN architecture employed in this study, in which the library of computer vision was imported, is depicted in Figure 2. The generator received noise-free and noisy images as inputs, and we attempted to use the discriminator to reconstruct images that were similar to noise-free data in the epochs. The epoch number was 200, but due to the small sizes of the metal implants, this network was unable to reconstruct implants. The generator and discriminator networks consisted of UNet and a convolutional network, respectively. Sixty-four kernel filters of 4×4 in size with Leaky ReLU activation, batch normalization with the same padding, upsampling 2D, convolutional 2D transpose with the concatenating layers and loss function of binary cross-entropy were used as generator structure while using similar information as in the first part except using transpose convolutional 2D as discriminator structure were the architecture of the GAN network.[27]

Quality image metrics

All images were compared using image quality measurements. The measurements included normalized root-mean-square error (NRMSE), peak of signal-to-noise ratio (PSNR), contrast-to-noise ratio (CNR), and structural similarity index between two images (SSIM). These metrics were calculated using a Google Colaboratory notebook by writing codes for estimating the quality of whole images.[28] To compare quality image indices, noise-free images were compared to noisy images. Noise-free images were also compared to denoized images to observe each network's success in denoizing. The values for mean and standard deviation (std) in CNR, PSNR, SSIM, and NRMSE for all codes were calculated to compare networks. If high mean values and lower std values were obtained in calculating CNR, PSNR, and SSIM and lower mean and std values were derived with the NRMSE metric, this indicated that the image quality obtained by that neural network was the best.

Patient study

To produce training and validation data for the GAN, 40 CT images of patients with head and neck cancer were acquired on a syngo CT VC40 (Siemens, Shanghai Medical Equipment Ltd.) with slice thickness 1 mm and an in-plane resolution of 512×512 pixel. Image of dental

area slices were then segmented using RadiAnt DICOM software (version 5.0.1). The images were then exported to MATLAB software. The matrix of images was loaded in MATLAB and one or two dental implants were added to random locations of the segmented parts of the images. Dental implant densities added to matrix were equal to 1.3 times the number of teeth as depicted in Figure 3. Histogram equalization code was used in Python 3.7 to change the gray levels of all images to $+128$ from -128 which is vital that all images have the same gray levels and pixels.[29] In total, 8000 noisy and normal teeth images were generated for training in addition to the 2000 noisy and noise-free images for GAN network validation. This generation was done by "Image Data Generator" class by importing Keras library in python for image classification, this class will help us to train and generate more data for the GAN network with 450 epochs. Forty CT images of the patients were finally used for GAN network testing.

Comparisons were made of images before and after denoizing using three and four Region of interest (ROI) of images with one and two dental implants, respectively. ROI of images were located in center, the buccal area near metal implants and oral cavity, which is shown in Figure 4. These regions were selected with the same X and Y axes and

areas (155 mm^2) in noisy and denoized images using ImageJ software (version 1.52a). Using this software, we estimated the max, min, mean, and std of gray levels. The formulas to calculate these quality image metrics are as Eq. 2:

$$CNR = (max-min)/std$$

$$PSNR = 10 \log (max^2/std)$$

$$NRMSE = \sqrt{(max^2 + min^2)}/CNR \quad (2)$$

The process flows for training, validating, and testing the network are shown in Figure 5. Calculating image improvement after denoizing for each network is shown in the Eq. 3. A, B, and C denote the differences in image indices between noisy and denoized images. The improvement was defined using the average A, B, and C values. Numbers 1 and 2 refer to image indices between noise-free and noisy images and between noise-free and denoized images, respectively.

$$Mean\ A = (CNR2 – CNR1)/CNR1$$

$$Mean\ B = (PSNR2 – PSNR1)/PSNR1$$

$$Mean\ C = (SSIM2 – SSIM1)/SSIM1$$

$$Improvement\ for\ each\ Network = ([A + B + C]/3) \times 100 \quad (3)$$
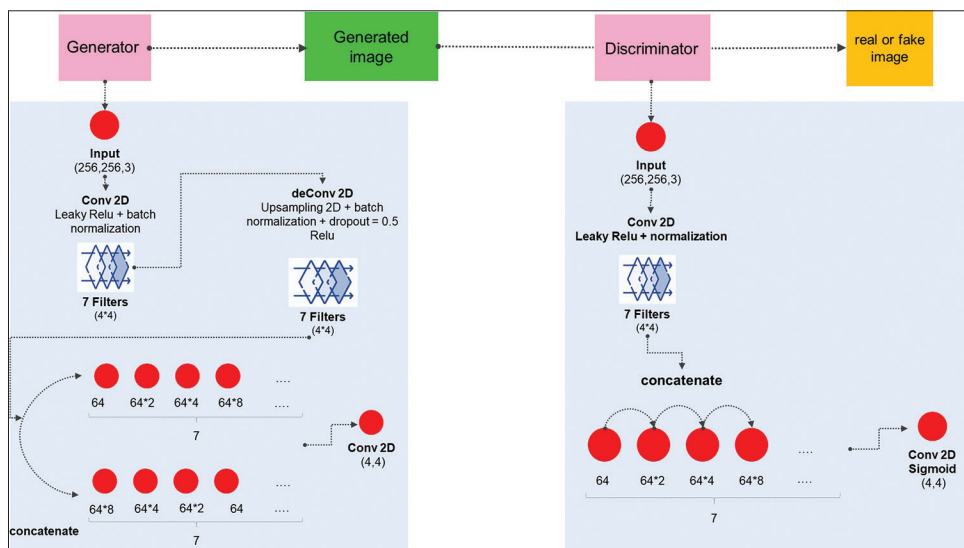

Figure 2: Generative adversarial neural networks architectural flow, with generator and discriminator architecture of network shown separately in blue rectangles
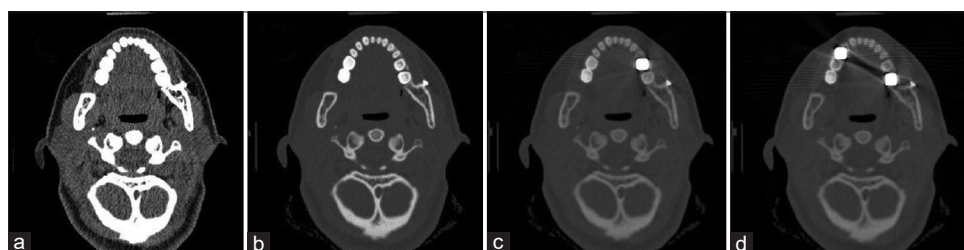

Figure 3: Images for training the generative adversarial neural network. (a) Original image, (b) Histogram equalization on original image, (c) simulated image with one dental implant, (d) simulated image with two dental implants used for GAN training and validation of input data

Results

Phantom study results

Figure 6 shows test images after the number of metal artifacts was reduced using DnCNN, UNet, and GAN. Figures S1 and S2 show loss and accuracy diagrams for training and validation, respectively, according to the network shown. For all networks except for the GAN, Figures 6 and S2 indicate that denoized images had less contrast and resolution as compared with images before denoizing. However, a comparison of image quality factors,
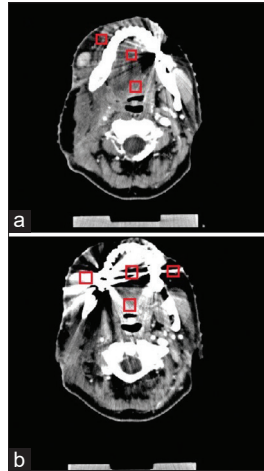

Figure 4: Location of ROIs in patient image; (a) With one dental implant, (b) With two dental implants

loss, and accuracy plots of networks [Figure S1], which were very low in terms of achieving high-quality images, showed that these five codes are not suitable for medical image processing and treatment. Still, GAN was more successful in denoizing images.

Table 1 shows image quality metrics between noise-free and noisy images. In addition, based on the image indices of CNR, PSNR, NRMSE, and SSIM for the six architectures following the denoizing of metal artifacts, denoized and noise-free images were compared, as shown in Table 1. A comparison of image indices as shown in Table 1 reveals that for the AE, UNet, and VGG architectures, only the CNR metric was improved after denoizing. For DnCNN and ResNet architecture, all image indices worsened. All GAN network metrics were improved after MAR, and the denoized images were very clear after noise reduction. While some previous studies showed the successful ability of networks in the MAR process, statistics related to image improvement as listed in Table 1 show that the GAN's improvement was 14%, whereas other network improvement was <zero.

Patient study results

CT images of patients were denoized and are depicted in Figure 7, which shows that image contrast was qualified. The quality indices of images were calculated between denoized and noisy patient images, and their improvements are shown in Table 2. A statistical analysis of Table 2
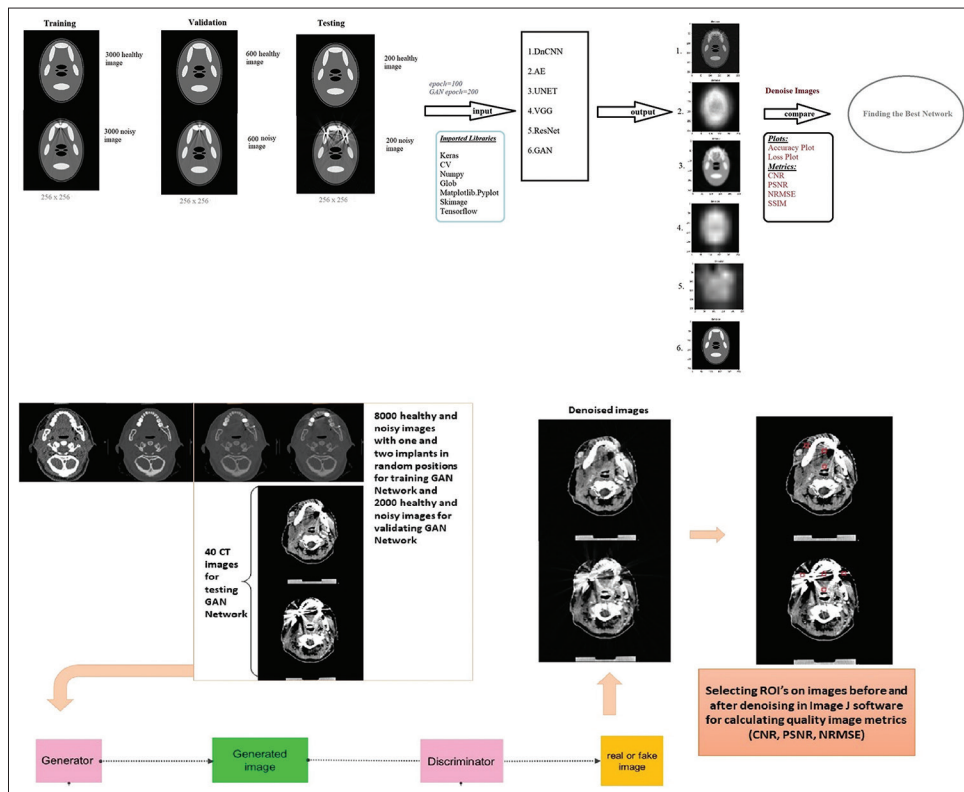

Figure 5: Flowchart of image reduction process with six types of neural networks and patient study
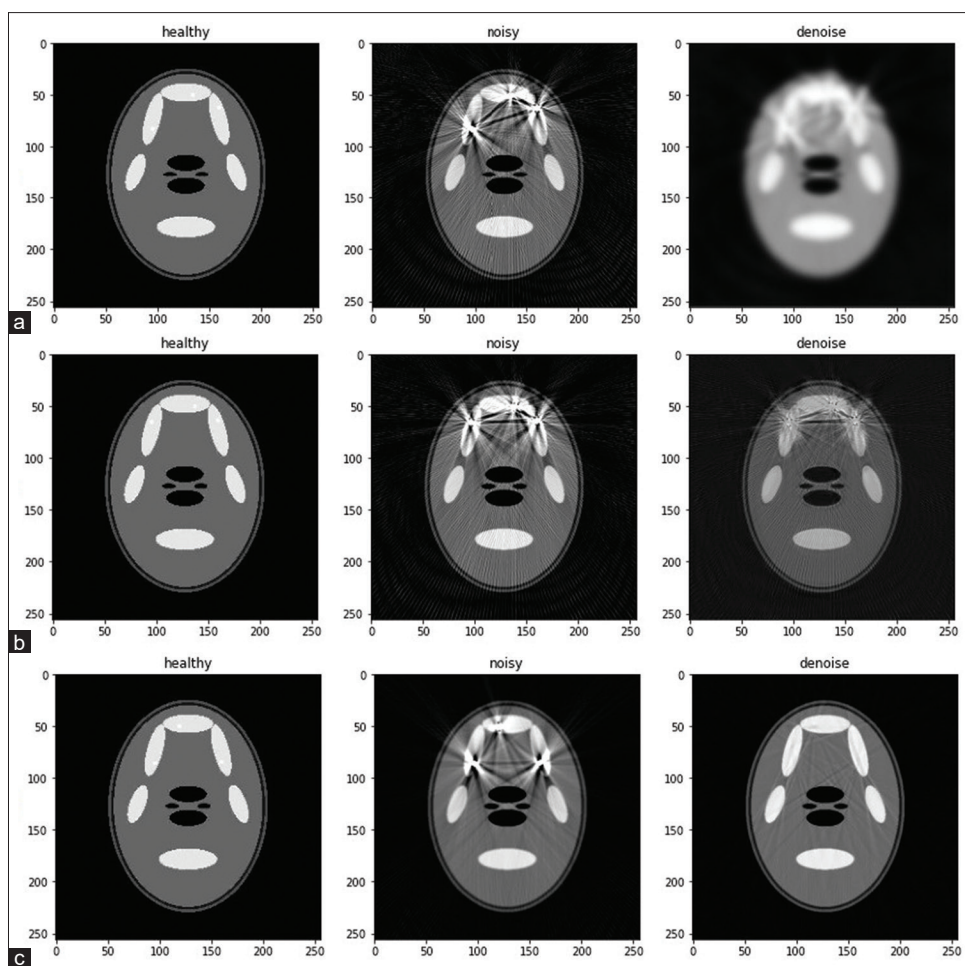
Figure 6: Noise-free, noisy, and denoised images obtained from popular networks. (a) UNet; (b) Denoising convolutional neural network; (c) Generative adversarial neural network

Table 1: Quality image metrics for six neural networks after denoising between noise-free and denoized images as compared with metrics before denoizing between noise-free and noisy images obtained from python

Networks	Mean (SD)				Improvement (%)
	CNR	PSNR	NRMSE	SSIM	
Comparison between noise-free and noisy images	0.96 (0.002)	23.39 (0.16)	0.22 (0.004)	0.81 (0.006)	
Comparison between noise-free and denoized images					
AE	1.29 (0.0004)	17.29 (0.01)	0.45 (0.0005)	0.52 (0.0009)	−7.78
UNet	1.16 (0.006)	15.12 (0.004)	0.58 (0.0003)	0.26 (0.0001)	−28.21
DnCNN	0.62 (0.02)	23.26 (0.1)	0.23 (0.002)	0.64 (0.01)	−18.85
ResNet	0.81 (0.01)	15.26 (0.4)	0.58 (0.03)	0.41 (0.02)	−33.03
VGG-16	1.22 (0.009)	15.32 (0.01)	0.57 (0.0008)	0.39 (0.0003)	−19.76
GAN	1.01 (0.01)	29.09 (0.1)	0.12 (0.002)	0.91 (0.005)	14

SD - Standard deviation; CNR - Contrast-to-noise ratio; PSNR - Peak of signal-to-noise ratio; NRMSE - Normalized root-mean-square error; SSIM - Structural similarity index measure; AE - Auto encoder; DnCNN - Denoizing convolutional neural network; ResNet - Residual network; VGG - Visual geometry group; GAN - Generative adversarial neural

showed that CT images of patients with one and two dental implants were improved in all ROI. Our results revealed that the centers of images showing a single dental implant were improved by 12.87% in terms of quality image metrics between noisy and denoized images. In the right buccal space (when the implant was on the left side of the patient's mouth), there was a 14.45% improvement in terms of image quality indices between noisy and denoized images. In the oral cavity, a 37.32% improvement in terms of image quality metrics was obtained, which appear near strong streaks in the images as a result of dental implants, and image improvement was more evident compared to

Table 2: Comparison of quality image metrics between noisy and denoized images of patients with one and two dental implants

ROI	Image	Mean	Minimum	Maximum	SD	CNR	PSNR	NRMSE	Improvement (%)
Image with one implant									
Center	Noisy	160.97	110	208	22.05	4.44	2.43	111.6	12.87
	D-noise	166.09	128	199	15.78	4.49	3.34	111.55	
Oral cavity	Noisy	114.49	59	164	32.01	3.28	1.53	96.23	37.32
	D-noise	134.85	96	179	18.73	4.43	2.71	96.48	
Right - buccal	Noisy	64.58	0	139	38.99	3.56	1.17	73.62	14.45
	D-noise	61.57	9	115	29.26	3.62	1.45	60.60	
Image with two implants									
Center	Noisy	215.72	162	255	23.11	4.02	2.50	150.58	2.11
	D-noise	207.94	164	243	20.58	3.8	2.76	149.63	
Oral cavity	Noisy	223.22	48	255	59.85	3.45	0.96	139.52	16.44
	D-noise	222.42	111	255	39.91	3.60	1.45	146.42	
Right - buccal	Noisy	113.99	0	253	86.52	2.92	0.66	147.95	84.55
	D-noise	93.47	48	150	25.98	3.92	1.82	79.48	
Left - buccal	Noisy	124.95	0	255	107.01	2.38	0.54	165.19	84.75
	D-noise	76.76	18	162	37.24	3.86	1.31	82.88	

SD - Standard deviation; CNR - Contrast-to-noise ratio; PSNR - Peak of signal-to-noise ratio; NRMSE - Normalized root-mean-square error; ROI - Region of interest
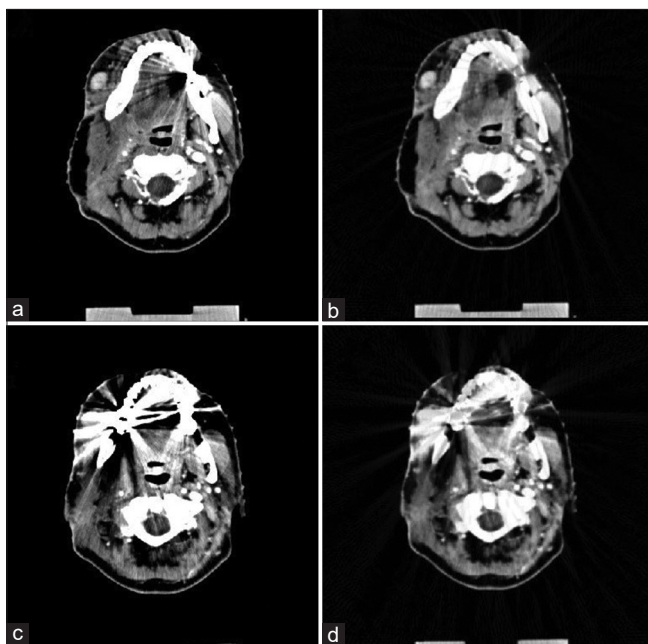

Figure 7: Computed tomography images of a single patient. (a) Noisy computed tomography image with one implant, (b) Denoized computed tomography image with one implant, (c) noisy computed tomography image with two dental implants, and (d) denoized computed tomography image with two dental implants

parts that were remote from the implanted tooth. However, in CT images with two dental implants (when the implants were on the left and right sides of the mouths of patients), the improvement was greater near dental implants and less in other ROI. More than 84.5% improvement could be observed near dental implants, whereas improvement in the oral cavity and center of the image was less than those images of mouths with single dental implants. The GAN showed its high ability to denoize artifacts derived from dental implants, which are very small and have less density than high atomic number (Z) materials.[29] Results from images comparisons showed that the GAN network successfully denoized and qualified images with powerful artifacts caused by metal prostheses, particularly near metal areas. This improvement showed that the GAN network is successful near strong streaks and can reduce metal artifacts near high Z materials, especially in unclear images with more artifacts.

Discussion

Researchers investigated that the Cycle-GAN can produce CT images that have realistic artifacts, which may provide a method of data augmentation. In Cycle GAN architecture, the PatchGAN discriminator only penalizes the generator at the scale of patches.[30] Considering Cycle GAN, we used UNet as the architecture of the generator, but we changed the discriminator architecture to CNN instead of using PatchGAN. On the other hand, we increased epochs number to 450 instead of 200 epochs to construct noisy images like healthy images.

Two multi-layer CNN architectures for denoizing low-dose CT images were surveyed: ResFCN and ResUNet. Training images were derived from realistic simulations using the XCAT phantom. The ResUNet approach indicated a PSNR of 44.00 as compared with 41.79 for ResFCN.[31] Using ResNet in our study showed less improvement in the PSNR metric in comparison to the GAN network.

A MAR method was proposed in another study[32] 3D adversarial nets were constructed using a regularized loss function designed for metal artifacts caused by multiple dental fillings. The suggested framework had an outstanding capacity for reducing robust artifacts and recovering the

underlying missing voxels. To overcome the limitations of noise reduction using voxel-wise regression, researchers in 2017 introduced[33] a noise-reducing generator CNN together with an adversarial discriminator CNN to develop a GAN. Results indicated that training with voxel-wise loss led to the highest PSNR with respect to referenced routine-dose images. GAN training improves the ability of CNNs to produce images with similar appearances to reference CT images with routine doses. This achievement helps us to use the CNN framework as GAN's discriminator architecture.

An effective method is suggested for reducing speckle noise and preserving texture details by the GAN. Several repeated densely sampled B-scan OCT images were used in another study[13] that employed multi-frame registration for denoizing generator training. Frequency-based error, PSNR, and SSIM were compared using DCSRN, GAN, UNet, and SRResNet networks, with results yielding image metrics of 3.63, 27.81, and 0.90, respectively, which were achieved using a GAN architecture and we used GAN as a successful network for our patient study.

A comparison of quality image metrics between six types of neural networks on simulated CT images of head-and-neck phantoms made this study different from others that used two to four networks with two or three image metrics. Most previous studies have employed CNN algorithms and popular MAR algorithms. However, developed architectures for use in studying medical images were evaluated. GAN was developed as a successful network, particularly for denoizing dental artifacts near images of implants with high streaks based on a comparison of image quality indices between noisy and denoized images. The use of a GAN as a new approach was affected by the numbers and positions of dental implants.

Conclusion

Results showed that artifacts could be denoized using the GAN network, and genuine CT images showed >84% improvement in images with two dental implants in the buccal and lateral areas. In addition, an improvement of >37% was achieved for images with single dental implants in the oral cavity area when image quality metrics were considered. These regions are important for head-and-neck cancer treatment during radiotherapy. The CT image process using a GAN network will help specialists to conduct accurate diagnoses of tumor positions and to help cure patients. It may also be used in examinations of other tumors that have high Z materials to reduce metal artifact effects and thus help cure people with different types of cancers.

Research ethics standards compliance

This research was conducted according to the declaration of principles for human studies. The protocol number of our ethics committee approval is IR.IUMS.REC.1397.231.

Informed consent was obtained in Farsi from patients who participated in clinical investigations.

Conflicts of interest

There are no conflicts of interest.

References

1. Barrett JF, Keat N. Artifacts in CT: Recognition and avoidance. Radiographics 2004;24:1679-91.
2. Park HS, Hwang D, Seo JK. Metal artifact reduction for polychromatic X-ray CT based on a beam-hardening corrector. IEEE Trans Med Imaging 2016;35:480-7.
3. Gjesteby L, De Man B, Jin Y, Paganetti H, Verburg J, Giantsoudi D, *et al.* Metal artifact reduction in CT: Where are we after four decades? IEEE Access 2016;4:5826-49.
4. Kim Y, Tomé WA, Bal M, McNutt TR, Spies L. The impact of dental metal artifacts on head and neck IMRT dose distributions. Radiother Oncol 2006;79:198-202.
5. Ziemann C, Stille M, Cremers F, Buzug TM, Rades D. Improvement of dose calculation in radiation therapy due to metal artifact correction using the augmented likelihood image reconstruction. J Appl Clin Med Phys 2018;19:227-33.
6. Men K, Zhang T, Chen X, Chen B, Tang Y, Wang S, *et al.* Fully automatic and robust segmentation of the clinical target volume for radiotherapy of breast cancer using big data and deep learning. Phys Med 2018;50:13-9.
7. Zhang D, Angel A. Single Energy Metal Artifact Reduction a Reliable Metal Management Tool in CT. White Paper. Canon Medical Systems; 2017.
8. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436-44.
9. Parenica HM, Ford JR, Mavroidis P, Li Y, Papanikolaou N, Stathakis S. Treatment planning dose accuracy improvement in the presence of dental implants. Med Dosim 2019;44:159-66.
10. Zhang Y, Yu H. Convolutional neural network based metal artifact reduction in X-ray computed tomography. IEEE Trans Med Imaging 2018;37:1370-81.
11. Gjesteby L, Yang Q, Xi Y, Shan H, Claus B, Jin Y, *et al*. Deep learning methods for CT image-domain metal artifact reduction. Developments in X-ray tomography XI. 2017;10391:103910W.
12. Wang J, Zhao Y, Noble JH, Dawant BM. Conditional generative adversarial networks for metal artifact reduction in CT images of the ear. Med Image Comput Comput Assist Interv

2018;11070:3-11.

13. Chen Z, Zeng Z, Shen H, Zheng X, Dai P, Ouyang P. DN-GAN: Denoising generative adversarial networks for speckle noise reduction in optical coherence tomography images. Biomed Signal Process Control 2020;55:101632.

14. Jain AK. Fundamentals of Digital Image Processing. New Jersey: Englewood Cliffs, Prentice Hall; 1989. p. 439.

15. Goncalves SB, Correia JH, Costa AC. Evaluation of dental implants using computed tomography. In: IEEE 3rd Portuguese Meeting in Bioengineering; 2013. p. 1-4.

16. Zhang K, Zuo W, Chen Y, Meng D, Zhang L. Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising. IEEE Trans Image Process 2017;26:3142-55.

17. Dertat A. Applied Deep Learning-Part 3: Autoencoders; 2017. Available from: https://medium.com/towards-data-science/applied-deep-learning-part-3-autoencoders-1c083af4d798.

18. Gondara L. Medical image denoising using convolutional denoising auto-encoders. In: IEEE 16th International Conference on Data Mining Workshops; 2016. p. 241-6.

19. Zhang J. UNet-Line by Line Explanation. Example UNET Implementation; 2019.

20. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention; 2015. p. 234-41.

21. Muneeb ul Hassan. VGG16 - Convolutional Network for Classification and Detection. In: Neurohive 2018.

22. Zhang X, Zou J, He K, Sun J. Accelerating very deep convolutional networks for classification and detection. IEEE Trans Pattern Anal Mach Intell 2016;38:1943-55.

23. Fung V. An Overview of Resnet and its Variants. Towards Data Science; 2017.

24. He K, Zhang X, Ren S, Sun J. Identity mappings in deep residual networks. In: European Conference on Computer Vision; 2016. p. 630-45.

25. Jian H. Pix2pix-Keras; 2019. Available from: https://github.com/He-Jian/pix2pix-keras/commit/43e479b56be3636dbf72192370bb8860cb1d0978.

26. Yi X, Walia E, Babyn P. Generative adversarial network in medical imaging: A review. Med Image Anal 2019;58:101552.

27. Goodfellow I. Nips 2016 Tutorial: Generative Adversarial Networks, arXiv No. 1701.00160; 2016.

28. van der Walt S, Schönberger JL, Nunez-Iglesias J, Boulogne F, Warner JD, Yager N, *et al.* scikit-image: Image processing in Python. PeerJ 2014;2:e453.

29. Khaleghi G, Hosntalab M, Sadeghi M, Reiazi R, Mahdavi SR. Metal artifact reduction in computed tomography images based on developed generative adversarial neural network. Inform Med Unlocked 2021;24:100573.

30. Du M, Liang K, Xing Y. Reduction of metal artefacts in CT with cycle-GAN. In: 2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC); 2018. p. 1-3.

31. Heinrich MP, Stille M, Buzug TM. Residual U-Net convolutional neural network architecture for low-dose CT denoising. Biomed Eng 2018;4:297-300.

32. Nakao M, Imanishi K, Ueda N, Imai Y, Kirita T, Matsuda T. Regularized three-dimensional generative adversarial nets for unsupervised metal artifact reduction in head and neck CT images. IEEE Access 2020;8:109453-65.

33. 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.
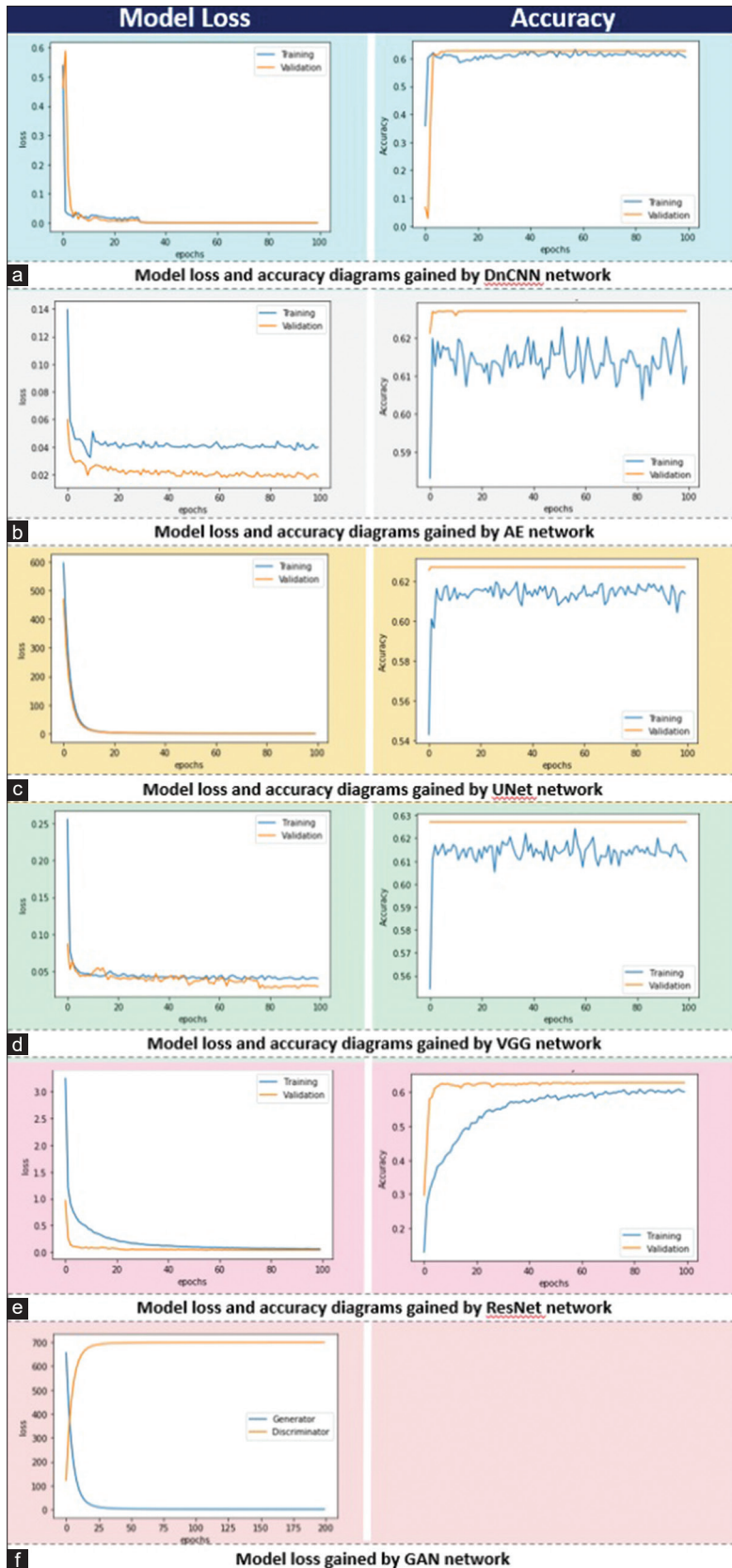
Figure S1: Loss and accuracy plots. Comparing Model Loss and Accuracy plots gained by Python 3.7 of all networks in Figure 1. Which shows that all networks were efficient in artifact reduction process. Model Loss and accuracy of networks gained by; (a) Denoizing convolutional neural network; (b) Auto encoder; (c) UNet; (d) visual geometry group; (e) Residual network; (f) generative adversarial neural network
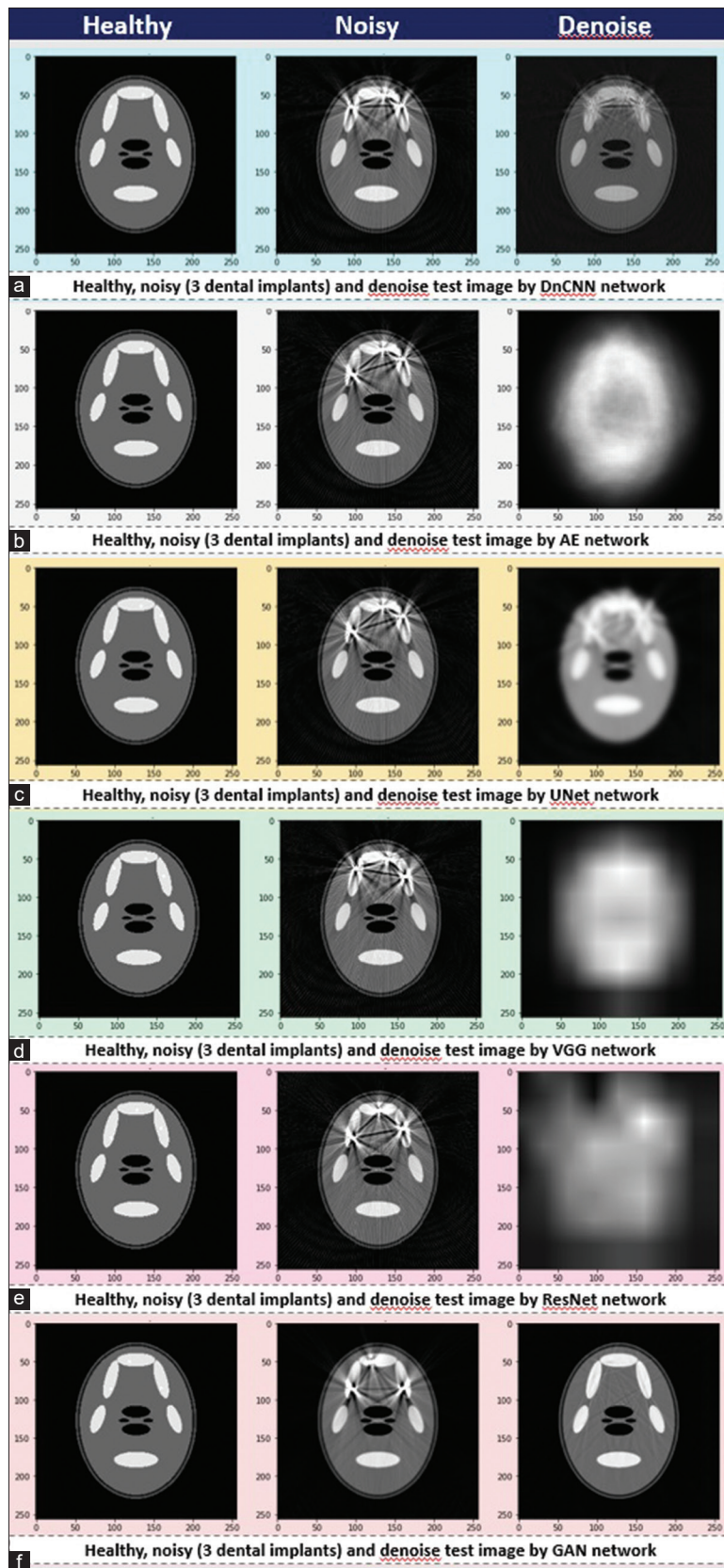
Figure S2: Denoised images. Comparing healthy, noisy and denoized images gained by Python 3.7 of all networks in Figure 2. Which shows that generative adversarial neural network was more successful in denoizing metal artifacts. Healthy, noisy, and denoized images gained by; (a) Denoizing convolutional neural network; (b) Auto encoder; (c) UNet; (d) Visual geometry group; (e) Residual network; (f) Generative adversarial neural network