Metal artifacts reduction in computed tomography by Fourier coefficient correction using convolutional neural network

Mai, Qi, Wan, Justin W.

Qi Mai, Justin W. L. Wan, "Metal artifacts reduction in computed tomography by Fourier coefficient correction using convolutional neural network," Proc. SPIE 11313, Medical Imaging 2020: Image Processing, 113132I (10 March 2020); doi: 10.1117/12.2549380

SPIE.

Event: SPIE Medical Imaging, 2020, Houston, Texas, United States

Metal Artifacts Reduction in Computed Tomography by Fourier Coefficient Correction using Convolutional Neural Network

Qi Mai[a] and Justin W.L. Wan[b]

[ab]University of Waterloo, 200 University Ave W, Waterloo, Canada

ABSTRACT

Metal artifacts are very common in CT scans since many patients have metal insertion or replacement to enhance functionality or mechanism of their bodies. These streaking artifacts could degrade CT image quality severely, and consequently, they could influence clinical diagnosis. In this paper, we propose to use the Fourier coefficients of a metal artifact-tainted image as the input to a convolutional neural network, and the Fourier coefficients of the corresponding clean image as target. We compare the performances of three convolutional neural network models with three kinds of inputs - sinograms with metal traces, images with streaks, and the Fourier coefficients of artifact-corrupted images. Using Fourier coefficients as inputs gives generally better artifacts reduction results in visualization and quantitative measures in different models.

Keywords: metal artifacts reduction (MAR), image restoration, Fourier Transform, convolutional neural network (CNN), computed tomography (CT)

1. INTRODUCTION

With the presence of metal implants in a patient's body, such as hip replacements and dental fillings (Figure 1), CT scans may contain different types of metal artifacts. Metal artifacts usually appear as dark or bright streaks, expanding from or surrounding around the metal pieces.[1] In the past few decades, plenty of metal reduction methods have been developed. However, so far, there is no standard solution to this difficult problem in clinical CT imaging.

Most of the existing metal artifacts reduction methods lie in three main categories: physical correction,[2] sinogram correction,[3] and iterative reconstruction.[4] Some approaches rely on the ensembles of the methods from different categories. However, many traditional methods still produce artifact residues in the reconstruction image or even introduce new artifacts. Convolutional Neural Network (CNN) has been widely used for computer vision and pattern recognition.[5] In recent years, it has also been applied to approach the problems in medical imaging. To tackle the metal artifacts reduction problem in CT scans, researchers usually take artifact-contaminated images for restoration or erroneous sinograms for inpainting. Zhang and Yu[6] proposed a two-phase CNN based method to suppress metal artifacts, by combining the information from the artifact-corrupted images and pre-corrected images obtained from linear interpolation in sinograms. Ghani and Karl proposed to let CNN learn the ways to inpaint the metal traces in sinograms.[7]

In this paper, we introduce a novel input, the outcome of Fourier Transform from an artifact-corrupted image. The corresponding target is the outcome of Fourier Transform from a clean image. Through passing the input to a CNN model, the frequencies of the signal corresponding to the streaks can be corrected. We tested the reconstruction result using three different CNN models with varied structures and found that using Fourier coefficients as input gives quantitatively and visually better results than using corrupted images or sinograms as input.

The rest of the report is organized as follows. Section 2 describes the Fourier Transform and the three model architectures. Section 3 introduces the method for data generation and training process, followed by results in

Further author information: (Send correspondence to Qi Mai)
Qi Mai: E-mail: qimai1618@gmail.com
Justin W.L. Wan: E-mail: justin.wan@uwaterloo.ca

Medical Imaging 2020: Image Processing, edited by Ivana Išgum, Bennett A. Landman,
Proc. of SPIE Vol. 11313, 113132I · © 2020 SPIE · CCC
code: 1605-7422/20/$21 · doi: 10.1117/12.2549380

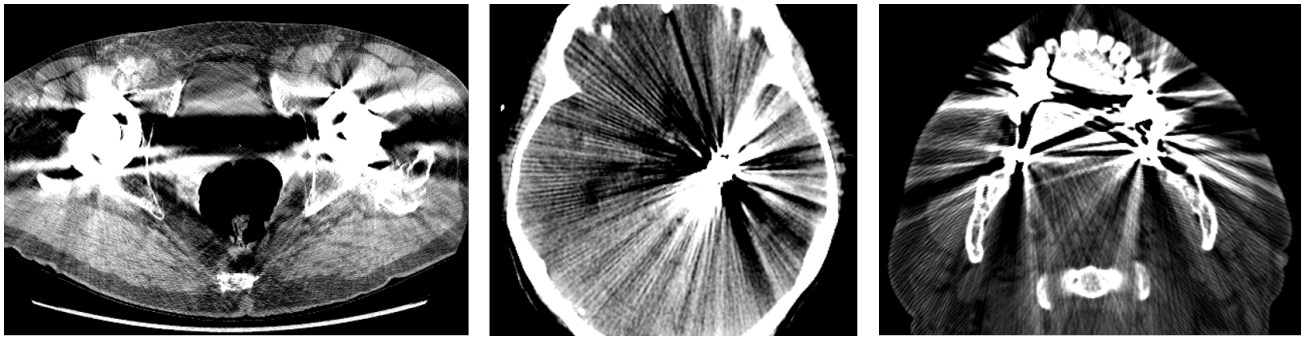

Figure 1. Examples of metal artifacts in the CT images of hip scan,[1] brain scan[8] and dental scan.[8]

Section 4. Lastly, we discuss the comparison of different inputs and models and suggest some potential direction for future work.

2. METHODOLOGY

2.1 Fourier Transform

In classical image processing, especially for medical images, Fourier Transform is a common and important analysis tool. By decomposing an image into sine and cosine components, we can obtain the representation of an image in Fourier domain.

Given a grayscale image $X = [f(x, y)]$, which is an $N \times N$ matrix, we can derive the frequencies of the signals in the image by Discrete Fourier Transform (DFT):

$$F(u, v) = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y) e^{\frac{-2\pi i}{N}(ux+vy)}$$

where $N \times N$ is image size. $F(u, v)$ consists of real and imaginary parts as $e^{i\theta} = \cos(\theta) + i\sin(\theta)$. In a similar way, we can recover the original image using inverse DFT:

$$f(x, y) = \frac{1}{N^2} \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} F(u, v) e^{\frac{2\pi i}{N}(ux+vy)}.$$

An image can be described as a signal, and thus, it can be expressed in terms of frequencies and magnitudes in Fourier domain. Usually, low frequencies capture the shapes in an image and high frequencies seize details such as edges and textures. Magnitude, $|F(u, v)|$, generally decreases as frequencies increase (Figure 2). So in the magnitude spectrum of an image, the central area usually looks brighter than its surroundings. Compared to the spectrum of a clean image, the spectrum of a corrupted image has more high frequencies with high magnitudes. Some bright rings spread out from the center are observable in the magnitude spectrum of the image with streaks, but absent in the spectrum of the clean image. These rings result from the metal and the streaks. By deriving the difference between the two spectra, the deviation appears at the rings and the off-center areas, since brighter areas indicate the bigger difference.

In the literature, to reduce artifacts using neural networks, it is normal to use either artifact-corrupted image or sinogram with metal traces as the input to the model.[7,9] We found out that using the Fourier coefficients as model input and target can be an alternate option. Normally, a model would learn a mapping from the image with streaks to the clean image, or from the sinogram with metal traces to the sinogram of the clean image. We propose that the input to a model is the Fourier coefficients of an artifact-corrupted image and the target is the Fourier coefficients of the corresponding clean image. Here, we expect that a model is able to learn a mapping in Fourier domain so that the Fourier coefficients corresponding to the streaks can be corrected.

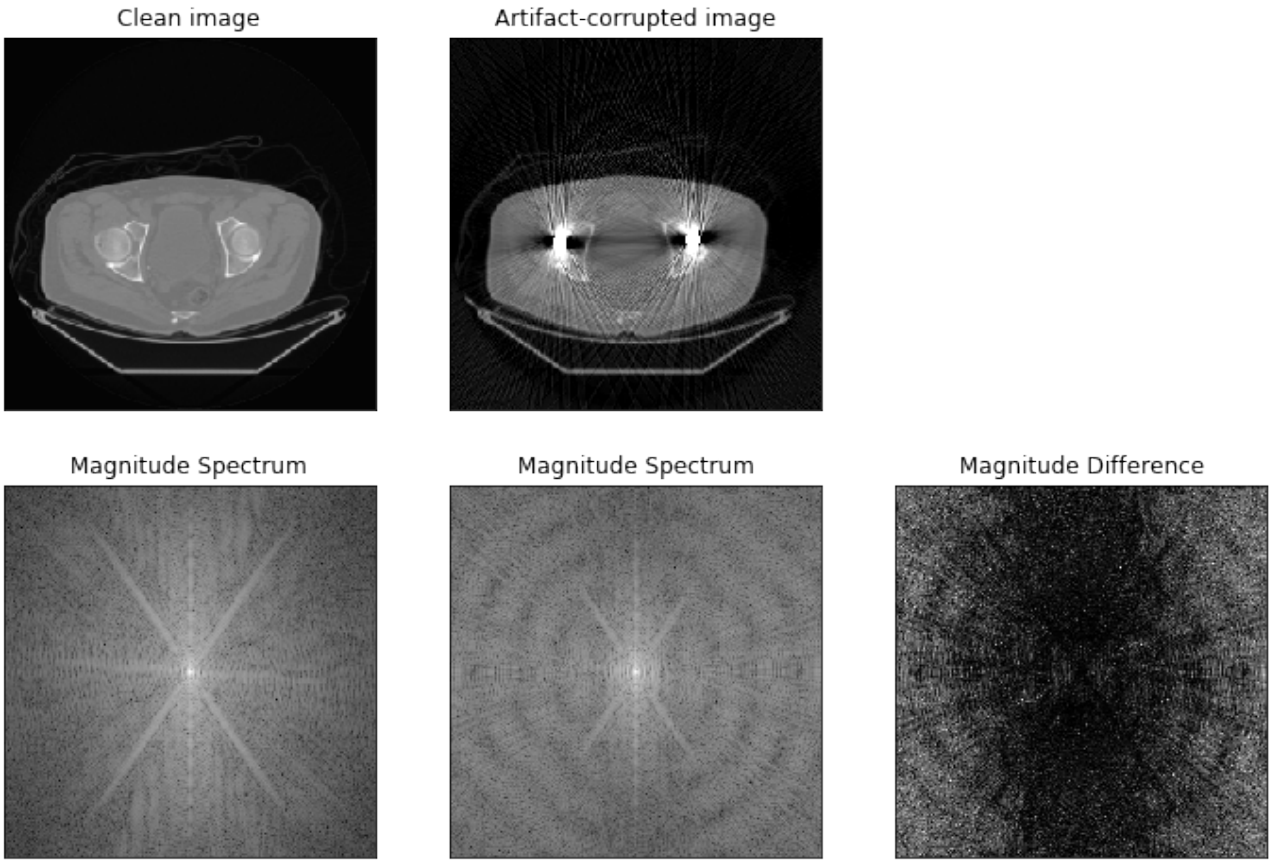

Figure 2. Magnitude spectra of clean image and artifact-corrupted image, and the absolute difference between the two spectra.

2.2 CNN Model Architectures

To see if our proposed model input, the Fourier coefficients of artifact-corrupted image, can be used for artifact reduction in CNN, we feed such input data to three CNN models with different architectures (Figure 3).

The first model is a Fully Convolutional Network (FCN) with 6 convolutional layers (FCN6). It is inspired by the model developed by Ghani and Karl,[7] who use such a model to interpolate metal traces in sinograms for metal artifacts reduction. Only convolutional layers are used in the network and the intermediate outputs from each layer are kept the same size. Also, we reduce the number of intermediate convolutional layers from 10, as proposed by Ghani and Karl, to 6 for better computational efficiency, as we find that there is no obvious degradation in model performance.

The second model, called SymNet, has 3 convolutional layers followed by 3 deconvolutional layers. The outputs from deconvolutional layers and convolutional layers are pair-wisely summed up and used as inputs to the subsequent layers. Specifically, (1) the input to the third deconvolutional layer is the sum of the output from the first convolutional layer and the output from the second deconvolutional layer; (2) the output of this model is the sum of the output from the second convolutional layer and the output from the third deconvolutional layer. This idea inherits from the model designed by Mao, Shen and Yang.[10] Adding the outputs of the symmetric convolutional layer and deconvolutional layer allows the network to learn the residuals instead of the direct mapping from input to target.

The third model, called UNet6, has max-pooling layers and upsampling layers which make it distinct from the other two models that keep the data size the same for each intermediate layer. Max pooling is applied after

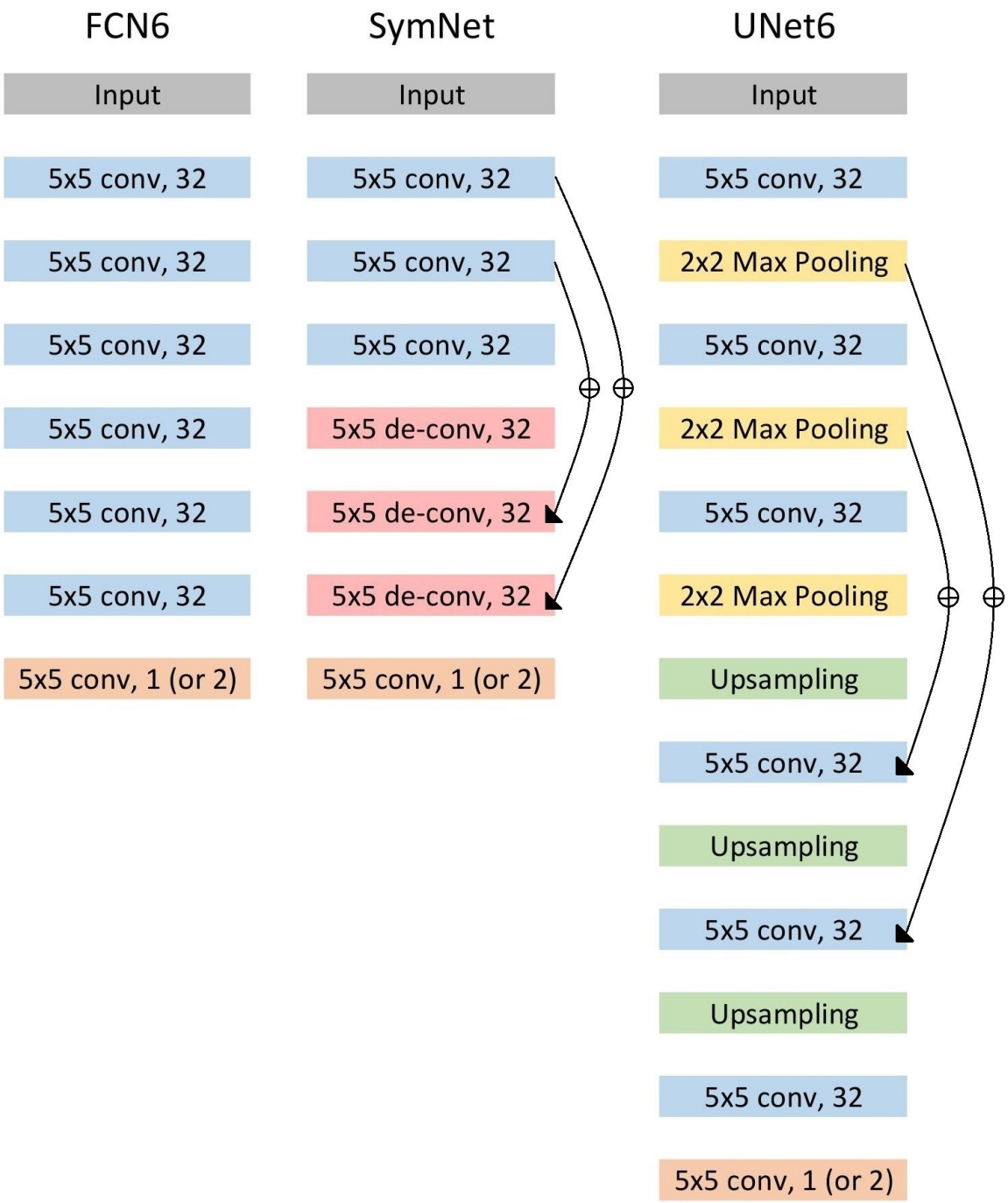

Figure 3. Three CNN model architectures. The first term in a cell specifies the type of a layer and the kernel size. The second term (if available) gives the number of kernel used in a layer. For example, 5×5 conv, 32 means 32 kernels with size 5×5 are used in a convolutional layer. Addition operations, which is used in SymNet and UNet6 for residual learning, are denoted by arrows and $\oplus$. The very last layer in each model is the output layer.

each of the first 3 convolutional layers, and upsampling is performed before each of the last 3 convolutional layers. Images are shrunk and then enlarged back so that the input size and output size remain the same. Additionally, we sum the output of the intermediate layers which have identical size for learning residuals. To be more specific, (1) the data used for the second upsampling is obtained from the sum of the output of the second pooling layer and the output of the fourth convolutional layer; (2) the data for the third upsampling is the sum of the output of the first pooling layer and the output of the fifth convolutional layer. As introduced by Ronneberger, Fischer and Brox,[11] the creators of U-net, the U-shape structure allows the network to extract meaningful features and attain precise localization.

In the models introduced above, the activation function, leaky Rectifier Linear Unit, and batch normalization are performed after each intermediate convolutional or deconvolutional layer.

3. EXPERIMENT

The image data we use for all experiments is obtained from Grossberg et al.[12] All the images in the dataset are metal-free and artifact-free*, and consequently, we simulate streaking artifacts to generate inputs. By applying FT on the clean images and those with simulated metal artifacts, we acquire their Fourier coefficients and treat them as targets and inputs respectively. We perform 9 experiments to compare the performance of FCN6, SymNet and UNet6, and the outcomes of different inputs - artifact-corrupted images, sinograms with metal traces and Fourier coefficients.

For data generation, we first select the slices of the hip part from the CT scans of a patient. For each of these slices (images), we manually insert two filled circles with proper radius, as two hip replacement metal pieces, at the centers of hip bones on both sides. By replacing the values in metal traces with the maximum value of the sinogram and using Filtered Back Projection (FBP), we acquire the images with simulated streak artifacts. As one patient has only around 10 slices of hip which are not enough to train a CNN model, we generate perturbations on the location and the radius of the inserted circles. With perturbations, we create 1100 images with metal artifacts, then separate them into a training set and a test set with 1000 and 100 images respectively.

As introduced in Section 2.1, after applying DFT on an image, we get a complex number, $a+bi$, corresponding to the frequency of a signal in the image. For each frequency, a and b are stored separately in two matrices, which are used as an input to the model. Since the models keep the size unchanged between input and output, we then apply inverse DFT to obtain the reconstructed images.

For each model we introduce in Section 2.2, we use an identical set of parameters for training. We deploy Adam optimizer with learning rate 5e-3 and decay 2.5e-5 and use 16 images in each batch. Training time varies due to different models and inputs, ranging from 1 to 2 hours. $60 \sim 90$ epochs are needed for convergence. The loss function is mean squared error, measuring the Euclidean distance between outputs and targets.

4. RESULT

Quantitative evaluations for reconstruction results are conducted using the following metrics.

- MSE: Mean Squared Error, the average squared difference between pixels. Smaller values imply smaller discrepancies between the artifact-reduced image and the clean image.
- PSNR: Peak Signal-to-Noise Ratio. Larger values indicate better reconstruction from the artifact-corrupted image.
- SSIM: Structural Similarity, ranged from 0 to 1. Higher values suggest higher similarity between the artifact-reduced image and the clean image.

*Unfortunately, we are unable to find clinical whole-body CT scans that have artifacts at the hip part, i.e. real world scans of the patients who have hip replacements. Such scans can make a huge contribution to this research in the future.

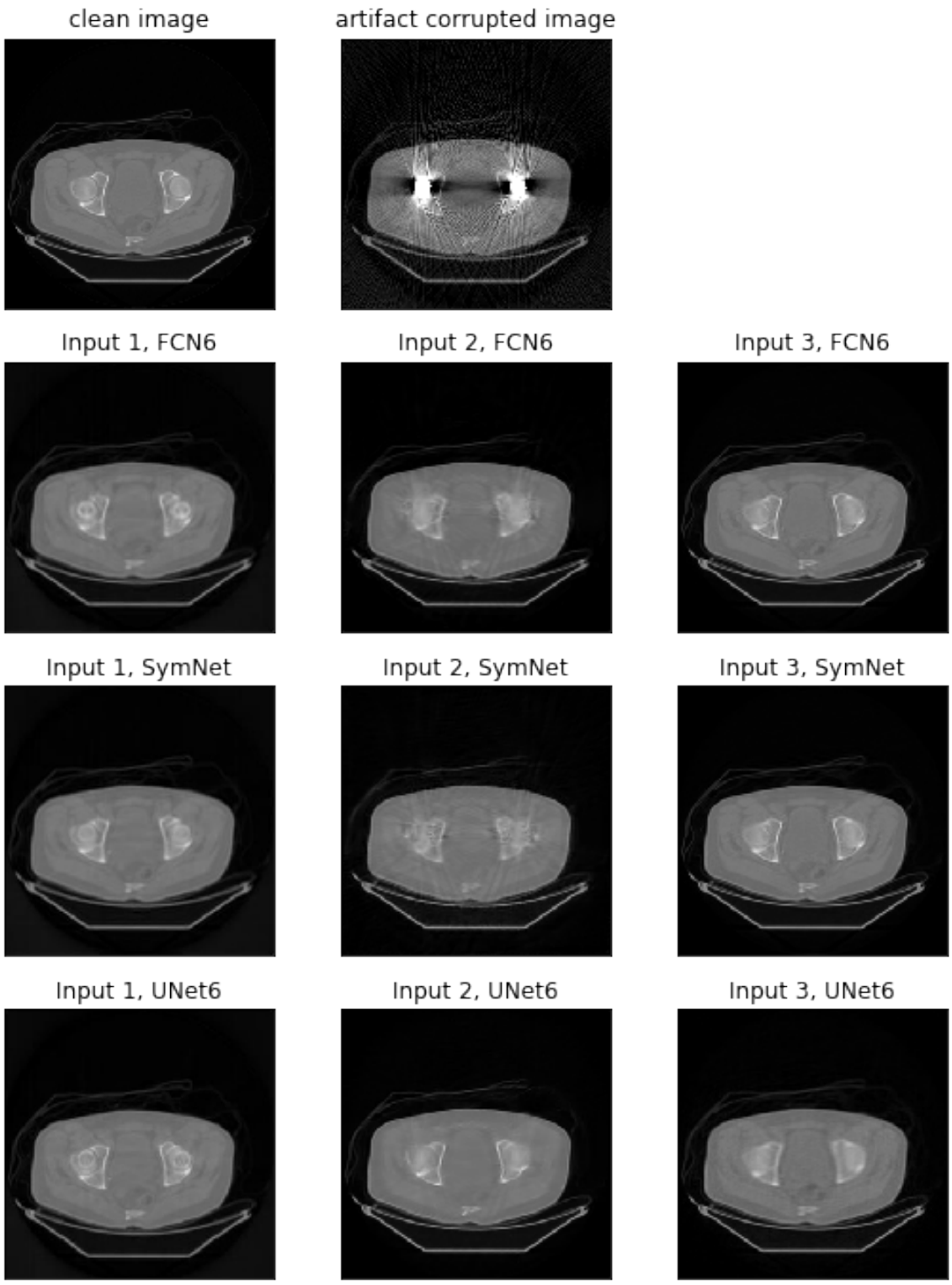

Figure 4. Metal artifacts reduction results of different inputs and different models. Input 1 is sinogram with metal traces, input 2 corresponds to using artifact-corrupted images as inputs, and input 3 means using Fourier Transform outcomes as inputs.

Table 1. Quantitative Results (test set)

Input	Model	MSE	PSNR	SSIM
Sinogram with metal traces	FCN6	9.180e-5	40.520	0.9752
	SymNet	**6.526e-5**	41.989	0.9777
	UNet6	9.969e-5	40.032	0.9746
Artifact-corrupted image	FCN6	5.633e-4	38.547	0.8823
	SymNet	5.963e-4	38.300	0.8938
	UNet6	2.130e-4	42.775	0.9684
Fourier coefficients of corrupted image	FCN6	9.909e-5	***46.096***	***0.9845***
	SymNet	9.943e-5	***46.081***	***0.9844***
	UNet6	2.712e-4	41.701	0.9621
Baseline		7.721e-3	27.221	0.5583

Most metal artifacts are suppressed by all models with any type of input, but the quality of the reconstructed images varies. We found out the reconstructed results through the correction in Fourier domain are visually (Figure 4) and quantitatively (Table 1) better than the ones through the correction in spatial domain and in sinogram. All three models perform similarly for sinograms, FCN6 and SymNet perform the best for Fourier coefficients, and UNet6 gives good results for artifact-corrupted images.

Results from correction in Fourier domain using FCN6 and SymNet preserve clear and sharp edges and visible background details. These models benefit from no pooling and learn the patterns in high frequencies for details and low frequencies for shapes. Frequencies contain little spatial relationship among the values, so maximum values cannot represent the information within the kernels. Thus, during training, lots of information are dropped in UNet6 but kept in FCN6 and SymNet. With UNet6, the reconstruction looks more blurred and its PSNR and SSIM are not as good as in FCN6 and SymNet. While taking artifact-corrupted images as inputs, we can still see some streaks remaining in the reconstructed images from FCN6 and SymNet. The details in the background and the edges of hip bones are blurred and hardly seen. Using images in spatial domain as inputs needs spatial feature extraction for artifacts reduction; thus, the results for these two networks are relatively poor. But spatial features can be extracted by pooling. Details are dropped after pooling layers but obtained back from upsampling. Therefore, UNet6 gives quantitatively and visually better results than the other two models, but still cannot compare with the best results from Fourier coefficients input.

Using sinograms with metal traces as inputs results in similar performance among these three models. SymNet, in this case, performs the best in terms of quantitative results, as it potentially keeps all detailed information and learns residuals which are essentially metal traces. This type of input leads to great results in terms of MSE. But the result cannot compare with the result of Fourier coefficients in terms of PSNR and SSIM. From the visual perspective, the reconstruction from FBP shows that some shades of streaks are present at the tissue between two hip bones.

From these experiments, we can see different neural network architectures have varied advantages and disadvantages, and they outperform each other when inputs take different representations of images. Usually, convolution takes the information of local neighbours and weights it using kernels. Since spatial information is collected during this process, the outcome tends to be blurred. Deconvolution is the reverse process of convolution and it is usually deployed to improve resolution. There is potentially a small amount of information lost in convolution and deconvolution, while a relatively large amount of data is lost in pooling, which takes advantage of strong features. Losing information is not desired when images in Fourier domain are used for inputs since the coefficients near the center are crucial for reconstruction. But for the images in spatial domain, a lot of spatial information can be extracted, shrunk and translated. Therefore, if Fourier coefficients as inputs are used, the model that can maintain the size of output to be identical in intermediate layers is preferred; if images with artifacts are used as inputs, the model that is good at spatial information extraction should be used.

5. DISCUSSION AND FUTURE WORK

In this paper, we introduce a new type of input to CNN models for metal artifacts reduction in CT scans, and compare the performance of three CNN architectures with three types of inputs. For visualization and quantitative measurements, in general, using Fourier coefficients as inputs produces better metal artifacts reduction results. When the input has a large amount of spatial information to be extracted and translated, UNet6, which is capable to draw important spatial features, is preferred. When we want all detailed information to be learned and nothing to be dropped out, consecutive convolutional or deconvolutional layers are preferred in a model without pooling layers.

Fourier Transform is an important tool in image processing, especially for medical images. Correcting Fourier coefficients for image restoration is a novel approach and the reconstruction results turn out to be better than using the image itself as input. With the consideration of the symmetric properties of Fourier Transform due to complex conjugation, $F(u,v) = \overline{F(-u,-v)}$ and $F(u,-v) = \overline{F(-u,v)}$, there are duplicated data stored in the two input matrices, which can be merged into one matrix.

Further, the clinical images we use for all experiments are clean, i.e. no artifacts in the scans. The metal pieces and artifacts are simulated to test our proposed method. In order to be clinically practical, it is desired to apply the method on the scans with real artifacts.

REFERENCES

[1] Boas, F. E. and Fleischmann, D., "CT artifacts: causes and reduction techniques," *Imaging Med* **4**(2), 229–240 (2012).

[2] Ketcham, R. A. and Hanna, R. D., "Beam hardening correction for x-ray computed tomography of heterogeneous natural materials," *Computers & geosciences* **67**, 49–61 (2014).

[3] Singh, S., Kalra, M. K., Hsieh, J., Licato, P. E., Do, S., Pien, H. H., and Blake, M. A., "Abdominal CT: comparison of adaptive statistical iterative and filtered back projection reconstruction techniques," *Radiology* **257**(2), 373–383 (2010).

[4] Wang, G., Snyder, D. L., O'Sullivan, J. A., and Vannier, M. W., "Iterative deblurring for CT metal artifact reduction," *IEEE transactions on medical imaging* **15**(5), 657–664 (1996).

[5] LeCun, Y., Bengio, Y., and Hinton, G., "Deep learning," *nature* **521**(7553), 436–444 (2015).

[6] Zhang, Y. and Yu, H., "Convolutional neural network based metal artifact reduction in x-ray computed tomography," *IEEE transactions on medical imaging* **37**(6), 1370–1381 (2018).

[7] Ghani, M. U. and Karl, W. C., "Deep learning based sinogram correction for metal artifact reduction," *Electronic Imaging* **2018**(15), 472–1 (2018).

[8] Radiology, R., "ReVision radiology: CT metal artifact reduction using the metal deletion technique (MDT)."

[9] Gjesteby, L., Shan, H., Yang, Q., Xi, Y., Claus, B., Jin, Y., De Man, B., and Wang, G., "Deep neural network for CT metal artifact reduction with a perceptual loss function," in [*In Proceedings of The Fifth International Conference on Image Formation in X-ray Computed Tomography*], (2018).

[10] Mao, X., Shen, C., and Yang, Y.-B., "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections," in [*Advances in neural information processing systems*], 2802–2810 (2016).

[11] Ronneberger, O., Fischer, P., and Brox, T., "U-net: Convolutional networks for biomedical image segmentation," in [*International Conference on Medical image computing and computer-assisted intervention*], 234–241, Springer (2015).

[12] Grossberg, A., Mohamed, A., Elhalawani, H., Bennett, W., Smith, K., Nolan, T., Chamchod, S., Kantor, M., Browne, T., Hutcheson, K., Gunn, G., Garden, A., Frank, S., Rosenthal, D., Freymann, J., and Fuller, C., "Data from head and neck cancer CT atlas," The Cancer Imaging Archive (2017).