Metal Artifacts Reduction in CT Scans using Convolutional Neural Network with Ground Truth Elimination

Qi Mai[1] and Justin W.L. Wan[2]

Abstract— **Metal artifacts are very common in CT scans since metal insertion or replacement is performed for enhancing certain functionality or mechanism of patient's body. These streak artifacts could degrade CT image quality severely, and consequently, they could influence clinician's diagnosis. Many existing supervised learning methods approaching this problem assume the availability of clean images data, images free of metal artifacts, at the part with metal implant. However, in clinical practices, those clean images do not usually exist. Therefore, there is no support for the existing supervised learning based methods to work clinically. We focus on reducing the streak artifacts on the hip scans and propose a convolutional neural network based method to eliminate the need of the clean images at the implant part during model training. The idea is to use the scans of the parts near the hip for model training. Our method is able to suppress the artifacts in corrupted images, highly improve the image quality, and preserve the details of surrounding tissues, without using any clean hip scans. We apply our method on clinical CT hip scans from multiple patients and obtain artifact-free images with high image quality.**

Keywords Computed Tomography (CT), Convolutional Neural Network (CNN), machine learning, ground truth

I. INTRODUCTION

When a patient has metallic objects implanted in their body, such as hip replacements, dental fillings, aneurysm clips and coils (Figure 1), their CT scans may contain different types of metal artifacts[1]. The typical appearance of metal artifacts are bright and dark streaks expanding from or surrounding the metal pieces. The streak artifacts obscure vital information for physicians to analyze CT scans and make diagnosis. In the past few decades, many metal artifacts reduction methods have been developed, but there is no standard solution to this difficult problem in clinical CT.

Conventional metal artifacts reduction methods usually leave artifacts in the reconstructed images or even create new artifacts, and many of them are not efficient[1]. In recent years, researchers have attempted to use deep learning to solve the problems in medical imaging. To tackle the metal artifacts reduction in CT scans, researchers often take artifact-contaminated images for restoration or erroneous sinograms for inpainting. Using a special loss function in a deep neural network, Gjesteby et al.[3] developed a model to suppress artifacts and tested it on phantom images. Ghani and Karl[2] inpaint sinograms using CNN. The methods in the literature are mostly based on supervised learning, which

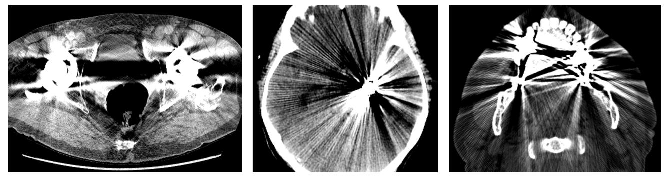

Fig. 1. Examples of metal artifacts in the CT images of hip scan[1], brain scan[5] and dental scan[5].

requires an input-target pair of data for each sample to train a model. Specifically, we need to have artifact-corrupted images as inputs and artifact-free images as targets for a CT scan-based model, or their corresponding sinograms as inputs and targets for a sinogram-based model. However, in clinical practices, the artifact-free scans at the part with inserted metallic objects are not available. Without artifact-free scans, which are usually call the ground truth images, model training cannot be completed.

In this paper, we focus on reducing metal artifacts in the hip scans of patients with hip prosthesis. In clinical practices, we do not have artifact-free hip scans as targets for a model to learn from. To solve this issue, we propose an innovative method that conducts model training on the scans near the hip with simulated artifacts. Artifacts suppression is then carried out on the actual artifact-corrupted hip scans using the trained model. Our approach eliminates the need for clean hip scans in model training and produces artifact-free hip scans using model prediction.

Our paper is organized as follows. Section II discusses the ground truth assumption issue and describes our proposed method. Section III presents different experimental results and Section IV points out future research directions.

II. METHODOLOGY

A. Our Proposed Method

Supervised learning requires input-target pairs for training. A model needs target examples to learn how to map an artifact-corrupted image to an artifact-free image, or how to map a sinogram with bright traces to a smooth sinogram without traces. But artifact-free scans do not exist in clinical practices. To tackle the ground truth assumption issue, we propose an innovative method to remove the need for clean images at the hip. In a series of CT scans, the scans from the abdomen to the thigh are similar in terms of body and bone shapes as well as tissues. By taking advantage of the similarities, our idea is to simulate artifacts in the scans near the hip and train a CNN model on the sinograms of these

[1]Qi Mai is with Computational Mathematics, Faculty of Mathematics, University of Waterloo, 200 University Ave W, Ontario, Canada. qimai1618@gmail.com

[2]Justin W.L. Wan is with David R. Cheriton School of Computer Science, Faculty of Mathematics, University of Waterloo, 200 University Ave W, Ontario, Canada. justin.wan@uwaterloo.ca

978-1-7281-1990-8/20/$31.00 ©2020 IEEE

images. After the model is trained, it can then be used to reduce artifacts in hip scans. In this situation, the training data, including the input images and the target images, is generated from the scans that are clinically available. The clean images of the artifact-corrupted scans at the hip are not required for model training any more.

To verify our idea, we prepare the training data as follows (see Figure 2). Given a series of artifact-corrupted hip scans, we first segment the metal pieces from these images. The segmentation result will be used to simulate artifacts on the scans near the hip. Since the trained model will be adopted to reduce artifacts on hip scans, the simulated artifacts are desired to be as similar as possible to the artifacts on the hip scans. Here, metal segmentation is used to approximate the shapes and locations of the metallic objects. We apply the K-means clustering method for metal pieces segmentation. After obtaining the segmented metallic objects, we store the results as masks for the subsequent streak simulation. We overlay a mask on a scan near the hip and apply radon transform to get a sinogram with bands in light colors. To simulate artifacts, we fill the bands with the largest existing value in a sinogram. Then Filtered Back Projection (FBP) is applied on the trace-filled sinograms to reconstruct CT scans. Due to the modification in sinograms, the reconstructed scans near the hip now have bright metal pieces as well as streak artifacts. Lastly, by using radon transform on the artifact-corrupted near-hip scans, we acquire their corresponding sinograms, which are the input data to our model. The targets are the sinograms of the clean images near the hip.

Training a neural network model often requires many samples since the model contains many parameters. The number of the abdomen scans and the thigh scans of a patient, which is usually around 20, is insufficient for a model to learn sinogram inpainting. We, therefore, conduct perturbations on the segmentation results regarding the sizes and locations of the metallic objects to obtain additional masks. In our experiments, 1000 masks are sufficient for model training. Then, we randomly choose a near hip scan and overlay one of the perturbed masks on it, to simulate artifacts. Through perturbation, we acquire sufficient training data and our model can learn to correct a variety of different traces in sinograms.

When our model finishes training, it can correct the sinograms by inpainting the bright bands. We use the trained model to correct the sinograms of artifact-corrupted hip scans. The corrected output sinograms will then be used to reconstruct artifact-free CT scans of the hip by FBP.

B. Single-Patient and Multiple-Patient Scenarios

Based on our proposed method, we design two model training processes for metal artifacts reduction in clinical applications. The first process trains models using the scans of a single patient and will be used on the hip scans of the same patient only. The other process trains model on the hip scans of multiple patients and is thus more robust. The trained model can be applied on the patients it trained on and potentially also applied on future patients.

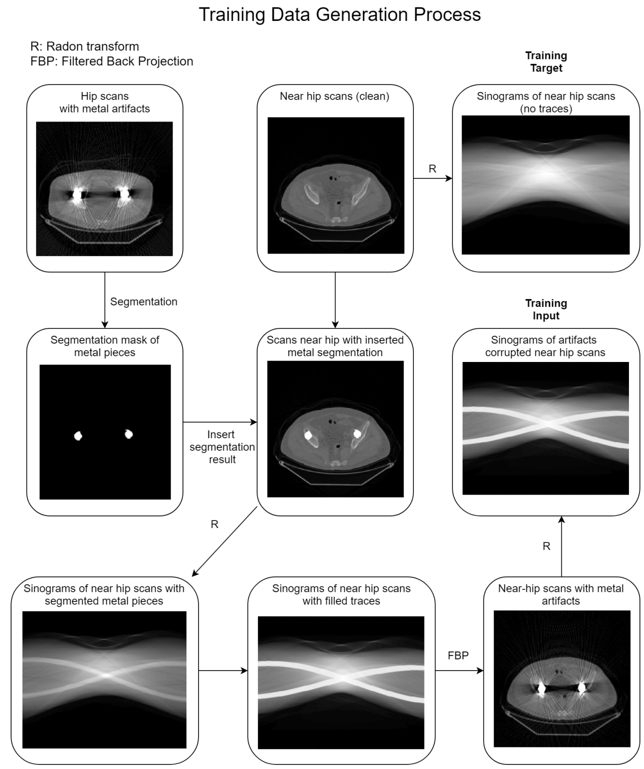

Fig. 2. Training data generation process. This process can be viewed as a segmentation step and an artifacts simulation process.

For the first process, we observe that consecutive slices from one CT screening are consistent in body shapes and similar in brightness, contrast and pixel value distributions. Their corresponding sinograms also share a large amount of resemblance. When a model is being trained on these sinograms, it does not need to adapt to background variations and thus, it can focus on inpainting traces. However, the generalizability and transferability of the trained model will be relatively low. The model may work effectively for the patient used for training, but may produce poor artifacts suppression results for other patients. Therefore, training a model on every single patient's sinograms is inefficient.

To resolve this concern, we consider developing a model that can be applied to multiple patients who have various physical circumstances and characteristics. Due to the diversity in body shapes and image properties, the input sinograms from different patients have more variances in pixel value distributions than the sinograms from a single patient. Unfortunately, our model trained on multiple patients fails to adapt to the dissimilarities in the inputs and produces sinograms that output blurred reconstructed CT images. Therefore, we consider a different approach by calibrating the inputs obtained from multiple patients. Among the CT scans of different patients, the inconsistencies fall in body shapes as well as the brightness and contrast of the scans. Since it is inappropriate for us to alter any physical property or condition of a patient, we ignore the body shape inconsistencies during input normalization. Between brightness and

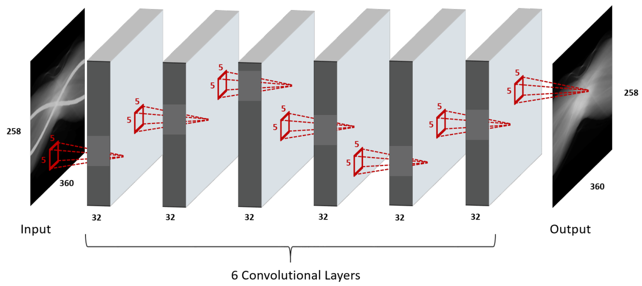

Fig. 3. Model architecture

contrast, we observe that brightness is the main cause of the variance in pixel value distributions of the scans and the corresponding sinograms. Thus, we focus on normalizing the brightness which can be done by scaling. Given an image S as the standard and an image I to be scaled, we find a scalar a such that the value distribution of aI is similar to the value distribution of S. We use the same scaling factor for all the scans of a single patient and different scaling factors for scans of different patients. After scaling the near hip scans of different patients and applying radon transform, we can acquire the sinograms with similar distributions. By applying FBP on the output sinograms from the model, we obtain artifact-free images with similar distributions. We then divide the images by their corresponding scaling factors to get the final images without artifacts, but with background pixel value distributions similar to the original scans.

C. CNN Model Architecture

The model we use is a CNN model (Figure 3) which inspired Ghani and Karl[2]'s work. There are 6 consecutive convolutional layers in the network. The kernels used in convolutional layers capture the spatial information in sinograms to correct the values of the bright traces. The kernel is of size 5×5 and is moved 1 pixel at a time. We pad zeros along margins so that the input image size is the same as the output image size. Each convolutional layer is followed by an activation function – Leaky Rectifier Linear Unit (Leaky ReLU) with slope 0.2. Additionally, batch normalization is performed after activation. The last convolution layer is the output layer and has no activation or batch normalization.

III. EXPERIMENTS AND RESULTS

A. Experiment Setup

The model we previously introduced is used for both experiments outlined in this section. We also apply identical training settings in all experiments. We deploy ADAM optimizer with learning rate 5e-3 and decay 2.5e-5. The batch size is 16 and the training time is 1.5 hours for 1000 sinograms. The loss function is mean squared error, measuring the Euclidean distance between outputs and targets.

The data used in the experiments are obtained from Grossberg et al.[4]. All the images in the dataset, including hip scans, are metal-free and artifact-free. Therefore, we need to simulate artifacts on hip scans and then validate our proposed method. We will use the clean hip images as reference images to perform both qualitative and quantitative evaluations. The evaluation will be conducted only on the test data in each experiment. Qualitative evaluations are performed by showing the artifact-corrupted images, the artifact-reduced images and the reference images. Quantitative evaluations are conducted using metrics MSE, SSIM and PSNR.

B. Experiment 1: single-patient scenario

In this experiment, we validate our idea by training our model on the sinograms generated from the near hip scans and testing on the sinograms of artifact-corrupted hip scans. All images in training and testing come from a single patient.

We generate 1000 hip scans with artifacts, then perform K-means segmentation with $k = 3$ to extract the masks of metallic objects. For generating training data, we pick 22 near hip scans, 11 consecutive slices above the hip and 11 consecutive slices below the hip from the same patient. We overlay each of the 1000 masks on a random near hip scan and then simulate artifacts as described in Figure 2. From this we acquire 1000 near hip scans with various artifacts. We use the corrupted sinograms of these 1000 scans as training input and the corresponding clean sinograms as training targets. The 11 slices with manually inserted circles are used to make 11 artifact-corrupted images, which are then used as the testing set. In model prediction, the 11 sinograms of the hip scans with artifacts will be used as inputs.

For experimental purposes only, we compare the reconstruction results of the 11 hip scans with the clean ones which are obtained from the data source directly. Clinically, we do not have the clean images of the hip to compare with. The only available comparison, in reality, is between artifact-corrupted scans and artifact-reduced scans.

As observed from Figure 4 and Table I, the artifacts are significantly suppressed in the hip scans transformed from the corrected sinograms. Our model successfully learns the way to correct the values in the traces regardless of different body parts. Even though the model has never seen the hip sinograms, by learning from the sinograms of abdomen and thigh, it is able to adapt the knowledge learned and fill the traces in the hip sinograms with appropriate values. This experiment demonstrates the flexibility of our model for handling discrepancies between training and testing sets.

TABLE I
QUANTITATIVE COMPARISON RESULTS OF EXPERIMENT 1

Images	MSE	SSIM	PSNR
Corrected Image	3.564e-4	40.990	0.9473
Uncorrected Image	6.623e-2	20.793	0.3514

C. Experiment 2: multiple-patient scenario

As explained in section II-B, building one model per patient is inefficient in practice. We hope to generalize our approach so that it can be used on multiple patients with various physical circumstances and characteristics.

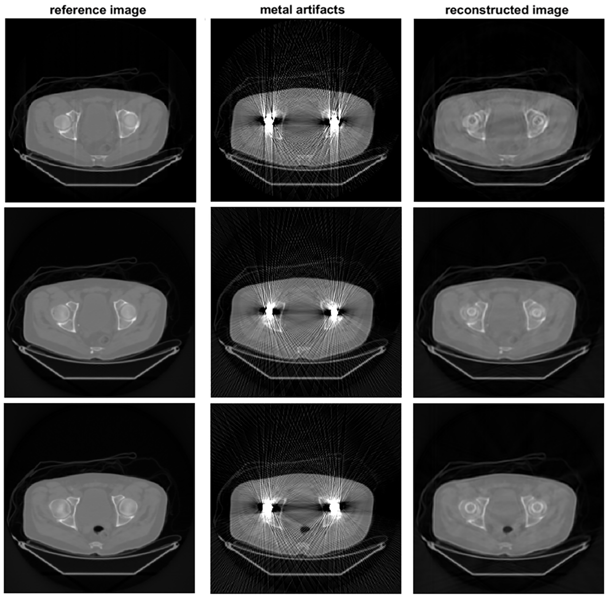

Fig. 4. Experiment 1: training on near hip scans and testing on hip scans from a single patient. Three testing examples to illustrate metal artifacts reduction result in the scans of hip.

In this experiment, we generate our training data set and testing data set using the scans from 10 different patients. We create 1000 corrupted images from the near hip slices from the 10 patients' scans for training using the same process in Experiment 1. The testing data in this experiment is 102 artifact-corrupted hip scans obtained from the 10 patients. Pixel value distribution normalization is conducted on these images before the input sinograms are generated. The reconstructed results of hip scans will then be re-scaled using the inverses of the scaling functions, and compared with the clean images in the same way as in Experiment 1.

We obtain the results shown in Figure 5 and Table II by normalizing the pixel value distributions. We see a greater reduction of artifacts when comparing the reference images and the reconstructed ones. With the normalization of pixel value distributions using scaling functions, our approach can adapt to the variances in body shapes and physical conditions at the hip, and be applied on multiple patients.

TABLE II
QUANTITATIVE COMPARISON RESULTS OF EXPERIMENT 2

Images	MSE	SSIM	PSNR
Corrected Image	2.324e-4	36.915	0.9094
Uncorrected Image	6.271e-2	21.438	0.4329

IV. FUTURE WORK

Our proposed method attains great performance in the above experiments. However, the research is limited to hip prosthesis as the method requires consecutive scans around the body part with metal insertion. The approach has the potential to succeed on artifact-corrupted head CT scans but might not be suitable for dental CT scans. In addition, acquiring clinical artifact-corrupted images would be helpful to validate the effectiveness of our approach.

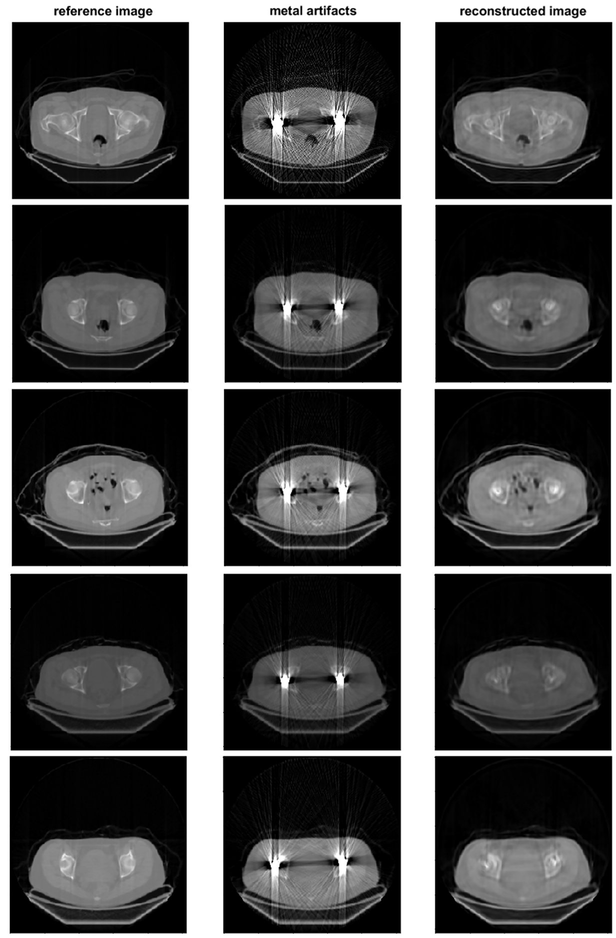

Fig. 5. Experiment 2: training on near hip scans and testing on hip scans of multiple patients. Five testing examples to illustrate metal artifacts reduction results in hip scans.

REFERENCES

[1] F Edward Boas and Dominik Fleischmann. "CT artifacts: causes and reduction techniques". In: *Imaging Med* 4.2 (2012), pp. 229–240.

[2] Muhammad Usman Ghani and W Clem Karl. "Deep learning based sinogram correction for metal artifact reduction". In: *Electronic Imaging* 2018.15 (2018), pp. 472–1.

[3] Lars Gjesteby et al. "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.

[4] A Grossberg et al. "Data from Head and Neck Cancer CT Atlas". In: The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.umz8dv6s.

[5] ReVision Radiology. *ReVision Radiology: CT metal artifact reduction using the Metal Deletion Technique (MDT)*. URL: http://www.revisionrads.com/index.html.