Sinogram + image domain neural network approach for metal artifact reduction in low-dose cone-beam computed tomography

Michael D. Ketcha,[a] Michael Marrama,[b] Andre Souza,[b] Ali Uneri,[a] Pengwei Wu,[a] Xiaoxuan Zhang,[a] Patrick A. Helm,[b] and Jeffrey H. Siewerdsen[a,*]

[a]Johns Hopkins University, Department of Biomedical Engineering, Baltimore Maryland, United States
[b]Medtronic, Littleton, Massachusetts, United States

Abstract

Purpose: Cone-beam computed tomography (CBCT) is commonly used in the operating room to evaluate the placement of surgical implants in relation to critical anatomical structures. A particularly problematic setting, however, is the imaging of metallic implants, where strong artifacts can obscure visualization of both the implant and surrounding anatomy. Such artifacts are compounded when combined with low-dose imaging techniques such as sparse-view acquisition.

Approach: This work presents a dual convolutional neural network approach, one operating in the sinogram domain and one in the reconstructed image domain, that is specifically designed for the physics and setting of intraoperative CBCT to address the sources of beam hardening and sparse view sampling that contribute to metal artifacts. The networks were trained with images from cadaver scans with simulated metal hardware.

Results: The trained networks were tested on images of cadavers with surgically implanted metal hardware, and performance was compared with a method operating in the image domain alone. While both methods removed most image artifacts, superior performance was observed for the dual-convolutional neural network (CNN) approach in which beam-hardening and view sampling effects were addressed in both the sinogram and image domain.

Conclusion: The work demonstrates an innovative approach for eliminating metal and sparsity artifacts in CBCT using a dual-CNN framework which does not require a metal segmentation.

[DOI: 10.1117/1.JMI.8.5.052103]

Keywords: metal artifact reduction; cone-beam computed tomography; low-dose imaging; image-guided surgery; spine surgery.

Paper 20276SSR received Oct. 16, 2020; accepted for publication Feb. 22, 2021; published online Mar. 13, 2021.

1 Introduction

Metal artifacts present a major challenge in cone-beam computed tomography (CBCT), often presenting as severe blooming, streaking, and shading artifacts that emanate from metal instrumentation and obscure critical anatomical structures. Such artifacts reduce the ability to accurately assess proper placement of metallic implants (e.g., pedicle screws in spine surgery), and they are compounded by sparse-view imaging techniques that are applied to reduce the radiation dose relative to full-view acquisitions (Fig. 1). The artifacts are caused by a combination of physical phenomena that, while always a concern, are amplified by the presence of metal. For instance, high-frequency metal edges can lead to streaking caused by motion or calibration

*Address all correspondence to Jeffrey H. Siewerdsen, jeff.siewerdsen@jhu.edu

2329-4302/2021/$28.00 © 2021 SPIE

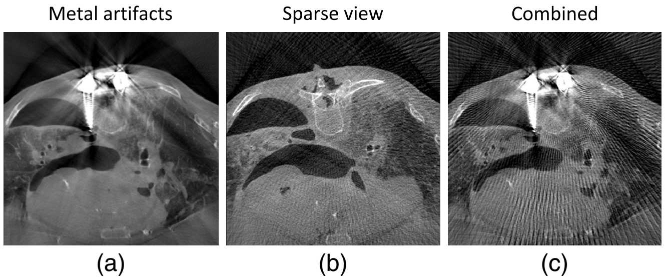

Fig. 1 Axial slices from CBCT reconstructions demonstrating: (a) metal artifacts in a standard full dose acquisition; (b) streak artifacts in a sparse-view acquisition (2-deg angular sampling over a half-scan acquisition); and (c) combined artifacts from metal implants and view sparsity.

errors,[1] and the strong attenuation of the metal results in regions of high scatter-to-primary ratio and photon starvation. Among the primary sources of these artifacts are two physical effects: (1) streaking caused by sparse-view acquisition (e.g., projection view sampling interval >1 deg); and (2) spectral beam hardening, which introduces non-linear bias in detector signal for rays traversing metal.

Traditional methods for metal artifact reduction (MAR) (summarized in Sec. 2.2) rely on techniques that are limited by the accuracy of metal segmentation (referred to as the metal trace) and tend to produce patchy image content while attempting to compensate for sparse-view acquisition. Fully image-based convolutional neural network (CNN) techniques have shown great promise in techniques to sharpen images[2] and reduce noise[3]; however, such techniques are not able to fully recover anatomical structures obscured by intense metal artifacts.

Motivated to address the factors of beam hardening and view sampling that underlie metal artifacts, we developed a CNN framework with two networks, each designed to address these sources of image quality degradation: (1) the first operates in the sinogram domain to mitigate biases directly in the projection data; and (2) the second operates in the image domain to reduce residual artifacts in the reconstructed image. Together, the framework addresses the streaks and shading associated with both undersampling effects in sparse data acquisition and x-ray spectral biases resulting from attenuation by metal.

2 Background

2.1 *Cone-Beam CT and Image Artifacts*

In CBCT-guided surgery, multiple scans may be acquired during a procedure, so minimizing the dose per scan is an important goal. One method to reduce dose is via sparse-view acquisition (using pulsed x-ray exposures) to acquire projection images at larger angular intervals (e.g., 2-deg intervals over 206 deg, as opposed to 1-deg intervals over 360 deg, amounting to a ∼2× reduction in view sampling density and a ∼3.5× reduction in dose), referred to below as a sparse short-scan. For standard filtered backprojection (FBP) reconstruction, such scan data result in images that are both noisy (∼1.9× increase in quantum noise) and subject to view sampling artifacts. For images acquired in the presence of metal, not only do standard beam-hardening metal artifacts arise [Fig. 1(a)], but strong view sampling streaks also occur [Fig. 1(c)] due to the high-frequency edges at the metal-tissue boundary.

The data associated with the sparse short-scan acquisition comprise 104 projection images, each 384×1024 pixels [$0.776 \times 0.388\ \mathrm{mm}^2$ pixel size, Fig. 2(b)]. As shown in Fig. 1, detector coordinates are denoted (u, v), and the number of projections is denoted p. Log-normalization and transpose of this stack of images yields the 384 sinograms used to reconstruct the 3D image [Fig. 2(b)], with FBP reconstruction yielding a $512 \times 512 \times 192$ volume ($0.415 \times 0.415 \times 0.83\ \mathrm{mm}^3$ voxel size). While such sinograms can be reconstructed sequentially in classic axial CT, forming one axial slice at a time, cone-beam geometry necessitates non-axial divergent backprojection in image reconstruction. CBCT backprojection is therefore performed over the entire reconstructed volume, as in the Feldkamp-Davis-Kress algorithm for 3D FBP.[4]

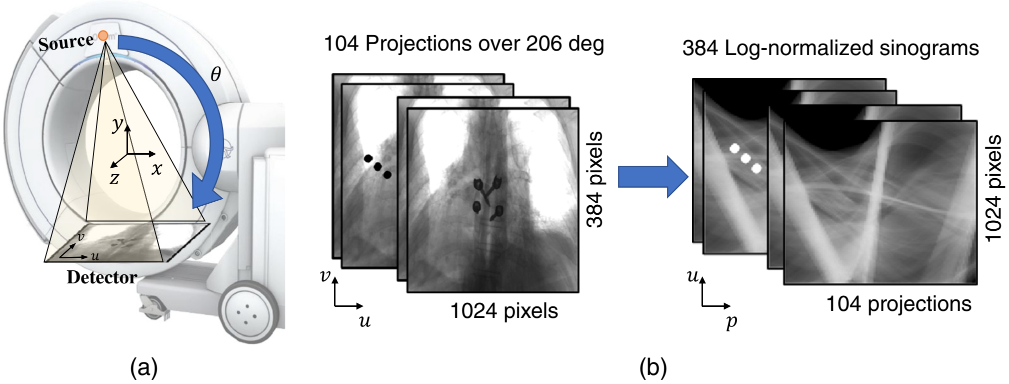

Fig. 2 Data for sparse-view image reconstruction. (a) Illustration of the O–arm™ imaging system showing cone-beam acquisition geometry. (b) Projection data acquired over a short-scan orbit are converted to log-normalized sinograms.

Due to the cone-beam geometry and fully 3D volumetric nature of the backprojection reconstruction, many existing methods for CNN-based MAR in CT (summarized in Sec. 2.3) are not readily transferable to CBCT, as they incur major memory limitations in processing the full 3D sinogram and image volume (rather than a single 2D sinogram and 2D axial slice).

2.2 *Traditional MAR and Sparse Reconstruction Techniques*

Traditional MAR methods commonly rely on segmentation of the metal in the reconstructed image volume (reconstructed without corrections), and the segmentation is forward-projected to achieve a sinogram metal trace that is in turn used to "inpaint" (i.e., replace with some computed value) the metal-containing regions of the original sinogram. A common approach uses linear interpolation for inpainting,[5] and more recent methods use an anatomical prior to normalize the sinogram prior to inpainting.[6]

A limiting step in traditional MAR is segmentation of the metal object in the uncorrected reconstruction. Because the initial image is contaminated with strong metal artifacts, accurate segmentation is difficult and often ad hoc because of inconsistencies in reconstructed voxel values and biases in the reconstructed CBCT image (e.g., due to scatter and truncation). Even small segmentation errors can result in strong artifacts in the 3D reconstruction: when a segmentation underestimates the extent of a metal object, the uncorrected pixel values result in residual artifacts; on the other hand, segmentations that overestimate the extent of a metal object cause inpainting of pixels corresponding to adjacent anatomy, blurring the region surrounding the metal object in an unrealistic manner. Moreover, such approaches are severely challenged when the metal component lies outside of the reconstructed field-of-view (FOV) and image-domain segmentation of the component is not achievable. Iterating on the segmentation estimate through successive passes of the approach described above can improve performance, but residual artifacts persist at a level that challenges clear visualization of metal and adjacent structures.

Such limitations may be addressed using model-based iterative reconstruction (MBIR) to reduce metal artifacts by down-weighting low-fidelity data and utilizing polyenergetic models for the beam and material attenuation during reconstruction.[7] Furthermore, known-component registration (KC-Reg)[8] and reconstruction (KC-Recon)[9] techniques have been developed to address such limitations, where a 3D model of the implant is registered to the projection data to obtain a high-fidelity metal trace that can be used for inpainting and/or MBIR. As an alternative to image-processing techniques, novel acquisition strategies have been developed that effectively avoid metal artifacts by means of tilted-axis[10] or non-circular[11] orbits to avoid projection angles that contain regions of severe beam hardening.

Sparse-view image reconstruction typically relies on sinogram interpolation or MBIR techniques. The former synthesizes higher angular sampling density in the sinogram, originally investigated using linear interpolation,[12] and with modern implementations modeling view-to-view interpolation according to the scan geometry.[13] On the other hand, iterative reconstruction techniques using total-variation regularization methods[14,15] may be employed to account for the sparsity of the data; however, the regularization parameters associated with such approaches must be carefully controlled (in a manner that is data-dependent) to avoid residual piecewise-constant artifacts in the image.

2.3 *Learning-Based MAR Techniques*

Learning-based MAR techniques have been developed and can be separated according to the domain in which they operate, e.g., the sinogram, the reconstructed image, or both. Methods from Gjesteby et al.[16] and Xu and Dang[17] exclusively operate in the reconstructed image domain to directly remove metal artifacts from the image in a manner amenable to post-processing without access to the scan data.

Other methods, including Ghani and Karl,[18] Park et al.,[19] and Liao et al.[20] operate in a manner similar to traditional MAR in that they use a metal segmentation to inpaint within the projection/sinogram domain and avoid artifacts in the first place. Zhang and Yu[21] utilized a CNN to achieve a reduced artifact prior image that is forward-projected and used to inpaint the original sinogram. Lin et al.[22] developed a dual-domain technique, employing a network in both the sinogram domain for inpainting and in the image domain to enforce artifact suppression and sinogram consistency. By incorporating FBP reconstruction within the dual-domain network architecture, they were able to create an end-to-end model to backpropagate the image-domain loss information into the sinogram inpainting network, which reduced the introduction of new artifacts due to inconsistent sinogram modifications. The approach was shown to work well for CT imaging (i.e., slice-by-slice reconstruction); however, the full volume reconstruction in CBCT is challenged by the larger GPU memory requirements for CNN techniques.

While image domain methods can be effective in reducing metal artifacts, they can be limited in severely affected regions in which there is insufficient data fidelity (e.g., low signal and high noise) and/or insufficient sampling (e.g., sparse-view acquisition). CNN inpainting techniques tend to perform better than image-domain methods at eliminating severe artifacts; however, as with their traditional counterparts, they rely on an accurate metal segmentation prior to corrections, which is particularly challenging in intraoperative CBCT imaging where metal may be outside of the reconstructed FOV.

3 Methods

3.1 *Proposed Framework*

While previously reported learning-based methods address MAR in fully sampled data (often CT, with the notable exception of Lin et al.[22] addressing CBCT), the work reported below demonstrates a learning-based MAR technique for sparse-view CBCT imaging. We approach this challenge in a physically motivated, dual-network framework (referred to as CNNMAR-2)—the first network operating in the sinogram domain, and the second in the reconstructed image domain. Together, the approach addresses physical factors associated with both sparse-view sampling and beam-hardening artifacts. Unlike conventional MAR, the proposed approach does not rely on an explicit metal segmentation and it is non-iterative (i.e., free from repeated forward and backprojection as in conventional beam-hardening correction methods).

Figure 3 shows the general form of the CNNMAR-2 framework. Rather than processing the entire stack of sinograms or the entire image volume with a 3D CNN, the 3D images are reshaped into slabs of three neighboring slices (i.e., 2D images with three channels) and processed sequentially (or in sequential batches). The output of the 2D CNN in each scenario is a single-channel 2D image, relating to the middle channel of the three-channel input. Projection data (following standard correction of detector gain, offset, and pixel defects) are stacked in order of angular sampling and transposed to create the sinograms (Fig. 2). The sinograms are processed by the first CNN with slabs determined by detector row order, and the output sinograms are restacked for FBP reconstruction. Similarly, after each slab of axial slices is processed by the second CNN, the images are restacked into the image volume. The sinogram domain network addresses both sparsity and beam hardening, and the resulting FBP reconstruction of the corrected sinogram

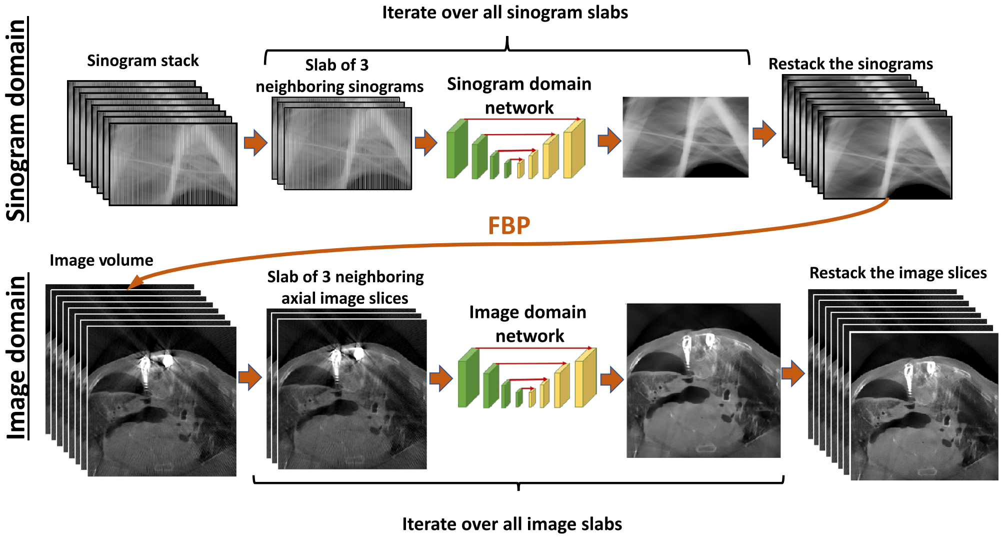

Fig. 3 Proposed framework (CNNMAR-2) for MAR in sparse-view CBCT. The first network operates in the sinogram domain and the second in the reconstructed image domain.

data is largely removed of biases, thereby leaving the image domain network to treat residual artifacts from imperfect or inconsistent corrections.

The approach is compared to an image-domain CNN (referred to as CNNMAR-1) that is tasked with removing both the sparse-view and beam-hardening artifacts in the reconstructed image. The approach is similar to prior work in MAR[16] and sparse-view data.[23] However, to our knowledge, such factors have not been simultaneously treated by means of an image-domain CNN. Both approaches are shown in Fig. 4(a). It is worthy to note that the input for the image-domain network for CNNMAR-1 is the sparsely reconstructed FBP (denoted "Sparse FBP"),

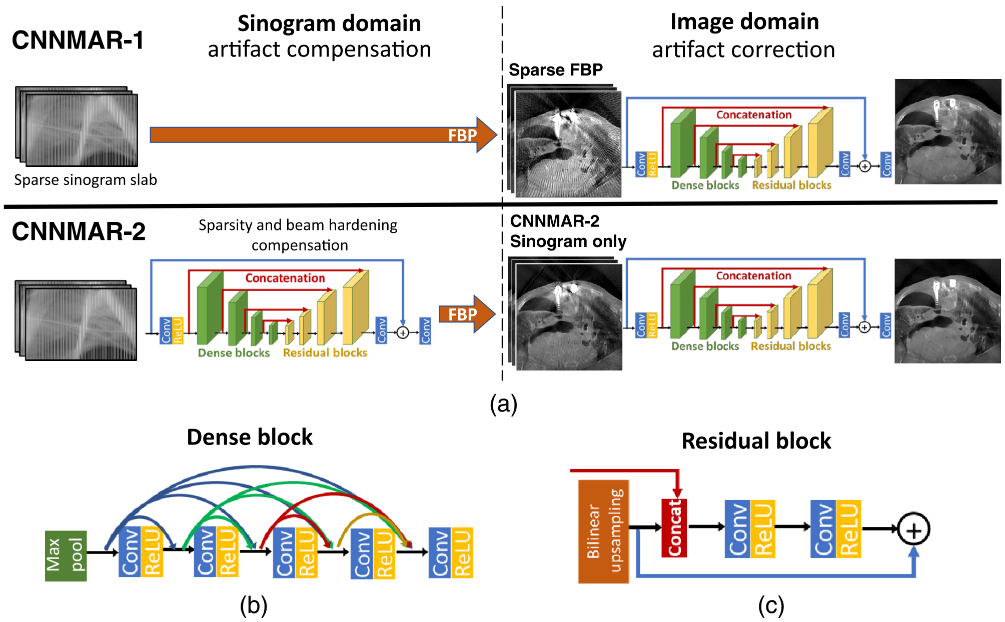

Fig. 4 (a) Flowcharts for implementations of the two CNN frameworks. The CNNMAR-1 method addresses both beam-hardening and sampling in the image domain following FBP reconstruction of the sparsely sampled acquisition. The CNNMAR-2 method compensates for both sparsity and beam-hardening artifacts in sinogram domain, while also removing residual artifacts in the image domain. Detailed structure of the (b) dense and (c) residual blocks.

whereas the input for the image-domain network of CNNMAR–2 is the FBP reconstruction resulting from the sinogram-domain network (denoted CNNMAR-2 Sinogram Only).

3.2 *Network Architecture*

The networks were implemented in TensorFlow[24] using the Keras API and trained with the Adam optimizer[25] with a learning rate of 10^{-4}. All networks used in this work used an identical architecture (Fig. 4), based generally on the architecture of Zhang,[23] which was designed for image-domain view-sampling artifact correction. The input slab is first processed by a 64-channel 7×7 ConvReLU, followed by four dense encoder blocks.[26] As shown in Fig. 4(b), the dense blocks begin with a 2×2 max-pooling, followed by four 64-channel 5×5 ConvReLU blocks (each with "dense" concatenation of the previous outputs), and finally a 64-channel 1×1 ConvReLU block. A series of four decoding residual blocks are then applied using a concatenation scheme similar to that of a U-Net.[27] The residual block [Fig. 4(c)] begins with 2×2 bilinear upsampling, and the output is concatenated with the associated encoder block output. The result is passed to a 128-channel, 5×5 ConvReLU block, and then a 64-channel, 1×1 ConvReLU block before being added to the output of the bilinear upsampling. The output of the decoder network is passed to a three-channel 3×3 Conv block, which is then added to the network input and passed to a final one-channel 7×7 Conv.

3.3 *Training Data and Metal Simulation*

Training data were generated from CBCT scans of multiple cadaver specimens imaged with the O-arm™ Imaging system (Medtronic, Minneapolis, Minnesota). The dataset comprised 25 thoracic and lumbar scans acquired over a 360-deg orbit—24 scans at 0.5-deg increments, and 1 scan at 1 deg increments. Six of the scans (three of which contained implanted spine pedicle screws) were reserved for testing. From the raw projection image data, sparse short-scans were generated by subsampling these scan data at 2-deg intervals over 206 deg.

Metal screw simulation was performed by augmenting the metal-free training data to include surgical screws. Computer-aided design models of spinal pedicle screws (each comprising a screw shaft and a polyaxial tulip head) of various lengths and diameters (31 shaft and two tulip head models) were randomly sampled and placed within the coordinate frame of the cadaver image. These screw models were then digitally forward-projected according to the projective geometry of the imaging system, yielding projection images that define the path lengths (in mm) for rays traveling through the metal component. As the shaft (s) and tulip (t) differ in material composition, they are forward-projected independently (yielding line integral distances $d_s(u, v)$ and $d_t(u, v)$ in mm for each projective angle) so that they may be treated separately for beam hardening simulations (Fig. 5, bottom left).

Prior to metal injection, a reduced tube current (mA) simulation[28] of the projection data $P(u, v)$ was performed to achieve physically accurate noise augmentation. Simulated screw projections were added to the existing raw projection images of anatomy (Fig. 5) using the following simulation model based on the Beer–Lambert law:

$$P_{\text{poly}}(u, v) = (P(u, v) - S)e^{(f_{\text{poly}}(d_s)+f_{\text{poly}}(d_t))(u,v)} + S, \tag{1}$$

$$P_{\text{mono}}(u, v) = P(u, v)e^{-(\mu_s d_s+\mu_t d_t)(u,v)}, \tag{2}$$

where S is a small scatter constant nominally set as the 1 percentile value of $P(u, v)/2$ as a robust method to achieve a scatter-to-primary ratio of 1 in the most heavily attenuated regions (appropriate given use of an anti-scatter grid on the O-arm™ system). The term serves to model the pre-scatter nature of attenuation and does not model the additional scatter contribution of the metal which was assumed negligible in this work. The $f_{\text{poly}}(d_i)$ term in Eq. (1) defines the polyenergetic attenuation of a material as a function of ray distance, which in Eq. (2) is simply the monoenergetic form of the Beer–Lambert law where μ_i is the monoenergetic attenuation coefficient of the respective material. As the specific value for the monoenergetic attenuation coefficients here serves primarily purposes of visualization (noting that large values may lead to photon

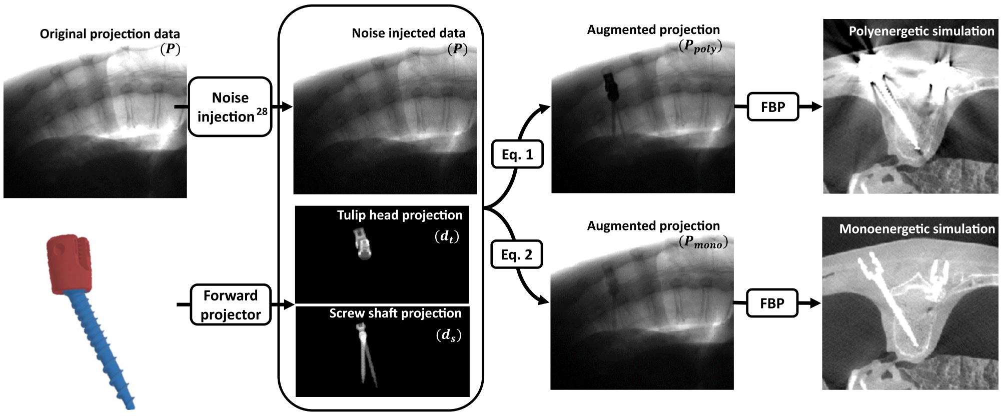

Fig. 5 Flowchart of the metal simulation process. Original projections were first optionally injected with noise to simulate low tube current projections. Then screws were defined in 3D anatomy and forward projected under the imaging system geometry of the projection images to obtain line-integral distances of the shafts (d_s) and tulips heads (d_t). These simulated forward projections were used to augment the metal-free projection data to create metal-injected projections P_{poly} and P_{mono}. The images were then reconstructed with FBP to yield the polyenergetic and monoenergetic reconstructions. The lack of metal artifacts in the monoenergetic model is noteworthy.

starvation), values of 0.04 and 0.045 mm^{-1} were used for the shaft and tulip head, respectively. For the polyenergetic case [Eq. (1)], the beam hardening model was computed for the shaft and tulip head based on their material composition using the SPEKTR Toolkit.[29] Factors comprising the energy-dependent attenuation, x-ray beam fluence, and detector response were considered for determining the polyenergetic attenuation described as

$$f_{\text{poly}}(d) = \log \frac{\sum_E D(E)q(E)\exp(-\mu(E)d)}{\sum_E D(E)q(E)}, \tag{3}$$

where $D(E)$ is the energy-dependent detector response curve, $q(E)$ is the normalized source beam spectrum, and $\mu(E)$ is the energy-dependent attenuation coefficient for the associated material (titanium for the shaft, and a 70%/30% mixture of cobalt and chrome for the tulip head). As image truncation prevents accurate estimation of the cadaver shape and size, $q(E)$ was attenuated by 32 cm of water to achieve a reasonable approximation of the beam spectrum. A look-up table of Eq. (3) was created for each material at distance intervals of 0.01 mm, and values were linearly interpolated from this table during polyenergetic simulation. Use of the look-up tables with Eqs. (1) and (3) yields a fast, simple approximation of metal attenuation. However, because beam hardening for the tulip and shaft materials are treated independently [rather than having a single $f_{\text{poly}}(d_s, d_t)$ function] a residual bias is expected in regions of the projection where the tulip and shaft overlap. Given the relatively small size of the metal objects used in this work, the residual bias is expected to be small (~2% to 4%).

3.3.1 *Sinogram domain training*

Training for the sinogram-domain network of the CNNMAR-2 method was performed using a 2 deg sampling P_{poly} sinogram as the input image and the 1° sampling P_{mono} sinogram as the label. Nine metal-free cadaver scans were used for training. For each scan, 30 repetitions of the noise and screw injection (Fig. 5) was performed—each containing 2 to 8 screws that were independently and randomly placed in the anatomical space (not necessarily restricted to the spine) to obtain a large dataset of corresponding P_{poly} and P_{mono} sinograms. Noise injection was performed on 50% of the repetitions with tube current uniformly distributed down to 1/8th of the original mA. The P_{poly} sinograms were sub-sampled into 104-projection (2 deg intervals) short-scan datasets. Zeros were interleaved in each column of the input sinogram (creating a 208-projection dataset), and the ground truth P_{mono} sinograms were sub-sampled into 208-projection sinograms with 1 deg intervals. Finally, the sinograms were log-normalized, and training was performed using the stack-of-neighboring slices method shown in Fig. 3 for 30 epochs with a batch size of 8.

The sinogram-domain network is not only tasked with upsampling the sinogram data, but also to compensate for metal-related biases by identifying (implicitly) metal regions and correcting the nonlinear polyenergetic attenuation with the linear monoenergetic attenuation model present in the ground-truth sinograms. Examination of our simulation model in Eqs. (1) and (2), following log-normalization by the initial fluence (I_0), shows that the input and ground truth images to the sinogram domain network can be written as

$$\mathrm{Sino}_{\mathrm{poly}} = -f_{\mathrm{poly}}(d_s) - f_{\mathrm{poly}}(d_t) - \log(P - S) + \log(I_0) - \log(1 + S/(P_{\mathrm{poly}} - S)), \tag{4}$$

$$\mathrm{Sino}_{\mathrm{mono}} = \mu_s d_s + \mu_t d_t - \log(P) + \log(I_0), \tag{5}$$

From Eqs. (4) and (5), we see the differences between the polyenergetic and monoenergetic training data (aside from sampling) arise from the polyenergetic $[f_{\mathrm{poly}}(d_i)]$ and monoenergetic $(\mu_i d_i)$ attenuation terms and the scatter term. Therefore, the proposed residual network is directly designed to linearize the polyenergetic $f_{\mathrm{poly}}(d_i)$ and reduce the effect of scatter at the metal regions while preserving the underlying anatomical data.

Based on the work of Tan et al.[30] in sinogram upsampling, the loss function utilizes an edge-weighted error term, which we augment by including an MS-SSIM[31] $[M(x,y)]$ term to help enforce structural similarity in the predicted sinogram. In total the term is

$$l_1(x,y) = \lambda_1(1 - M(x,y)) + (1 - \lambda_1)(\|x - y\|_2 + \alpha\|\nabla(y) \cdot (x - y)\|_2), \tag{6}$$

where ∇ is the gradient magnitude and the empirically determined weights tested on small training batches were $\lambda_1 = 0.83$ and $\alpha = 0.05$.

3.3.2 *Image domain training*

A separate dataset of 10 metal-free cadaver scans was used to train the image-domain methods of CNNMAR-1 and CNNMAR-2. Similar to Sec. 3.3.1, screw injection was repeated 30 times for each scan. The P_{poly} images were sub-sampled into 104-projection (2-deg intervals) half-scan images. For CNNMAR-1, the training input images were generated from FBP reconstructions of these sinograms. For CNNMAR-2, the sinograms were corrected and upsampled using the fully trained sinogram-domain network of Sec. 3.3.1 prior to FBP reconstruction. As the sinogram network quickly learned the simulation model of Eqs. (4) and (5), a data augmentation strategy to increase the presence of artifact in the reconstruction was incorporated by applying a random scale factor (between 0.8 and 1.2) to $f_{\mathrm{poly}}(d)$ (equivalent to scaling the material density) during polyenergetic metal simulation. Ground truth labels for training were generated from FBP reconstructions of the fully sampled (0.5-deg intervals) half-scan using the P_{mono} projections. Training was performed using the stack-of-neighboring slices method shown in Fig. 3 for 30 epochs with a batch size of 12.

Based on the work of Gjesteby et al.[16] and Yang et al.,[3] a perceptual content loss function[2] was used for its ability to remove noise while maintaining fine structure in the image. The perceptual content loss function minimizes the mean-squared error between a VGG-19 feature representation ($\phi(\cdot)$, output of the block5_conv4 layer, pretrained on ImageNet data) of the output (x) and ground-truth (y).

$$l_2(x,y) = \|\phi(x) - \phi(y)\|_2. \tag{7}$$

3.4 *Evaluation*

3.4.1 *Evaluation on sparse data containing metal*

Evaluation was performed on images of three cadaveric specimens with surgically implanted pedicle screws (not requiring metal simulation). Sparse-view acquisitions were sampled from complete scans, thus allowing comparison between the sparse-view and standard dose images. Both CNNMAR-1 and CNNMAR-2 were applied on the sparse-view acquisitions, and the resulting reconstructions were compared in terms of visual image quality, the fidelity of the anatomical structures near the metal objects, and quantitative differences in artifact magnitude adjacent to metal.

Qualitative visual inspection was performed to evaluate the reduction of beam-hardening and sparse-view artifacts while preserving the underlying anatomical structure. As the clinical context pertains to pedicle screw placement, emphasis was placed on the ability to detect screw breaches at the medial and lateral vertebral walls.

Given the importance of detecting breaches in a CBCT acquisition, the reduction of artifacts must not come at the cost of anatomical fidelity, e.g., by blurring out anatomical structure to remove artifacts. Quantification of this effect was performed by computing the standard deviation of the gradient magnitude image in manually defined regions-of-interest (ROIs)—a metric which is related to objective functions for auto-focus and CBCT motion compensation techniques.[32] The metric quantifies the assumption that stronger variation in the image gradient content implies sharper image content, and the metric would be reduced if a MAR method simply blurred out regions of heavy artifact without recovering the anatomy. To ensure the metric quantified only anatomical gradient structure (and not the metal artifacts themselves), seven ROIs covering 41 axial slices were selected between two screws (particularly emphasizing the spinal canal) where both CNNMAR methods successfully removed streaking artifacts, thereby leaving only relevant anatomy.

Quantification of artifact removal was performed by examining the ROI standard deviation in the regions connecting two tulip heads, which presents a region of particularly intense artifact with reliable positioning. Seven ROIs covering 32 axial slices, were selected in the regions between two tulip heads, which should otherwise have been homogenous image content. High standard deviation in these regions thus imply a strong presence of streaking artifacts.

3.4.2 *Evaluation on sparse data with simulated metal*

Evaluation was also performed on three cadaveric specimens with simulated metal data, where ground-truth anatomy was available. Following the polyenergetic simulation of pedicle screws according to Fig. 5 (without noise simulation), the scans were sub-sampled to 104 projection (2-deg intervals) sparse-view acquisitions and reconstructed using CNNMAR-1 and CNNMAR-2. Ground truth images were defined using fully sampled (0.5-deg intervals) half-scan reconstructions with the monoenergetic metal simulation model.

Quantitative evaluation of the CNNMAR methods was performed by examining the RMSE at regions near the pedicle screws. Nine ROIs covering 73 axial slices near pairs of pedicles screws were selected and the root-mean-squared error (RMSE) was computed between the CNNMAR outputs and the monoenergetic ground truth. Metal and air regions in the reconstruction were segmented and removed from the RMSE calculation to focus the metric on anatomical structures.

Blooming artifacts are commonly associated with reconstructed metal objects and increase the apparent size of the objects making evaluation of breaches difficult. The blooming of the metal structures following reconstruction was evaluated by examining the ratio of the measured screw diameter to the true screw diameter for both CNNMAR methods. Thirty-seven cross-sections over 12 screws were manually delineated, and the measured diameter was defined as the full width at 20% of the max intensity value.

3.4.3 *Evaluation on sparse data at reduced tube current*

While the sparse-view, short-scan acquisition reduces the dose to nearly one-quarter of a standard acquisition, further dose reduction can be achieved by reducing the tube current (mA) for each projection image. A low-dose simulation tool[28] was utilized to observe how CNNMAR-2 performs at very low dose conditions reducing the mA (and thus dose) down to 1/32nd of the standard tube current. The simulation was performed on one of the cadaveric specimens containing metal, and CNNMAR-2 method was evaluated with respect to visual image quality, quantitative evaluation of the noise magnitude (standard deviation in a homogenous region), and spatial resolution. Spatial resolution was quantified by the width of the edge-spread function (ESF) computed at a tissue edge, with line profiles fitted according to the following edge model:

$$e(x) = a - \frac{c}{2}\,\mathrm{erf}\left(\frac{x-r}{\sqrt{2}\sigma_{esf}}\right), \tag{8}$$

where σ_{esf} denotes the width of the ESF and a, c, and r are fitting parameters corresponding to the tissue intensity, contrast, and distance from edge.

4 Results

4.1 *Results on Sparse Data Containing Metal*

Figure 6 depicts the output for each of the methods on three cadaver images, each containing surgically implanted metal screws. For each cadaver, the first image shows the Full Dose FBP condition with no MAR method applied (206-deg scan at 0.5-deg increments for cadavers 1 to 2, 1-deg increments for cadaver 3). The second image shows the Sparse FBP reconstruction (206 deg scan at 2 deg increments). The third column shows the output of the sinogram-only step of CNNMAR-2 followed by the final outputs of CNNMAR-1 and CNNMAR-2.

The output of CNNMAR-1 (which takes the Sparse FBP as input) indicates that the fully image-domain approach provides significant image quality improvement, removing nearly all

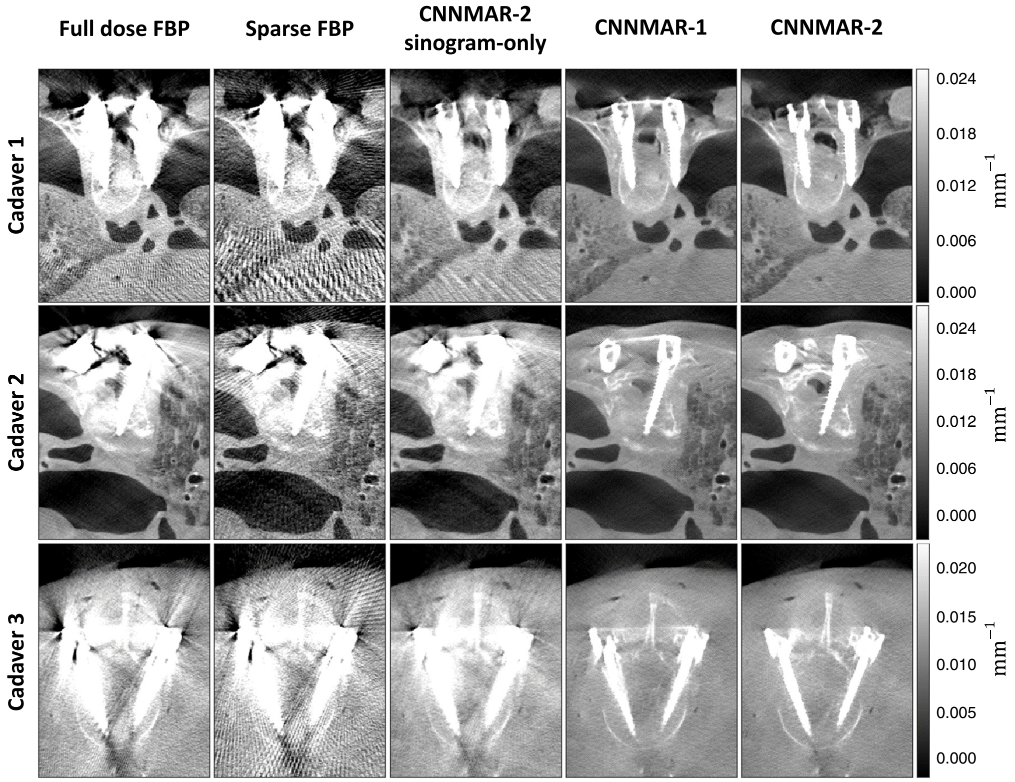

Fig. 6 Reconstructed images for three cadaver scans. From left-to-right: Full-Dose FBP, Sparse FBP, CNNMAR-2 Sinogram-Only, CNNMAR-1, and CNNMAR-2.

view-sampling artifacts and much of the beam-hardening artifact. Remaining artifacts appear localized to the tulip region (i.e., screw head), particularly in regions between two tulips. In such regions, it appears that the beam-hardening artifact almost completely obscured the anatomy, and the network learned to primarily blur out this region (as seen in Cadaver 2, CNNMAR-1). The CNNMAR-2 method addresses both view-sampling and beam-hardening artifacts prior to FBP reconstruction which is observed in Fig. 6 to remove nearly all the shadowing and streaking artifact. Not only is the artifact reduced, but the original anatomy (particularly in locations between two screws) is preserved. Much of this improvement can be attributed to the sinogram-domain step of CNNMAR–2, where, compared to the sparse FBP, the sinogram-only output significantly reduces both sparse-view and beam-hardening artifact, yielding a more reliable starting point for the image-domain network to correct the residual artifacts (compared the Sparse FBP starting point of CNNMAR-1).

Compared to standard inpainting methods, CNNMAR-2 does not rely on an explicit segmentation of the metal objects, typically generated by performing an uncorrected FBP reconstruction and applying a threshold segmentation. While such inpainting methods generally perform well at reducing metal artifacts, the performance is diminished when metal is present outside of the reconstructed FOV and a segmentation of the metal cannot be obtained, as observed in a different axial slice of Cadaver 1 [Fig. 7(b) which contains a pair of forceps outside the reconstructed FOV with artifacts emanating from the top-left corner of the image. The output of CNNMAR-2 [Fig. 7(c)], however, shows a robustness to the truncation of metal objects, which is a common scenario in the limited FOV setting of intraoperative CBCT.

Qualitative analysis in Fig. 6, indicated that while both CNNMAR methods significantly reduced metal artifacts, CNNMAR-1 tended to blur anatomical content in regions of strong artifact. The effect was quantified by computing the standard deviation of the gradient image in regions between the screws, particularly emphasizing the spinal canal. An exemplary region is shown in red boxes in Figs. 8(b) and 8(c), noting that regions were specifically selected to highlight pertinent anatomy while avoiding high-frequency artifacts. The box plots in Fig. 8(d) show a significant ($p < 0.005$, paired t-test) improvement in the standard deviation of the gradient within these regions, indicating that CNNMAR-2 better preserves anatomical content in regions of intense artifact.

Quantification of artifact removal was examined by computing the ROI standard deviation in the region between the tulip heads in regions that should otherwise be homogenous. Exemplary ROIs in Fig. 9 show such regions of consistent strong artifact due to the aligned metal edges of the tulip heads. While both CNNMAR methods remove the majority of the metal artifacts in the image (as observed in the qualitative analysis of Fig. 6), the boxplots in Fig. 9 show significant ($p < 0.005$, paired t-test) reduction of the artifacts in these particularly severe regions ($5 \times 10^{-3} \pm 2 \times 10^{-3}$ mm^{-1} [mean ± std] for CNNMAR–1 versus $3.5 \times 10^{-3} \pm 1.3 \times 10^{-3}$ mm^{-1} for CNNMAR-2).

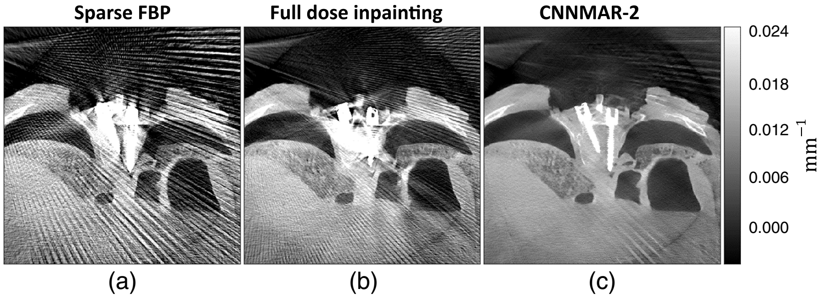

Fig. 7 Illustration of performance when metal is present outside the reconstructed FOV. In this case, a pair of forceps cause metal artifacts emanating from the top-left corner of the image. From left-to-right: (a) Sparse FBP; (b) Full Dose Reconstruction using a standard MAR inpainting method; and (c) CNNMAR-2.

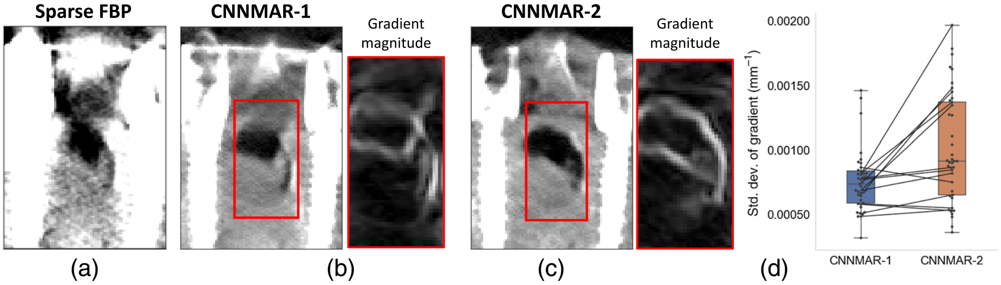

Fig. 8 Quantification of anatomical structure preservation. (a) Sparse FBP for reference. (b) CNNMAR-1 and (c) CNNMAR-2 outputs. (d) Standard deviation of the gradient magnitude image is evaluated region between two metal tulip heads that contain pertinent anatomical content (spinal canal) and avoid high-frequency artifacts [exemplary ROI shown in red boxes (b) and (c) with gradient magnitude image to the right]. Higher values indicate that more anatomy is preserved.

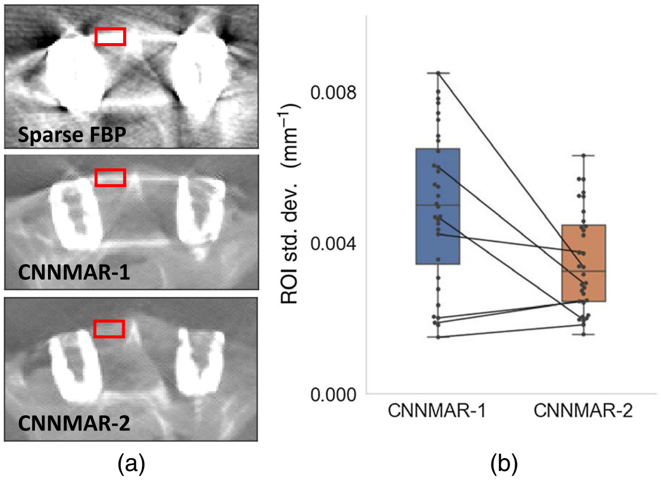

Fig. 9 Quantification of artifact removal. (a) Sparse FBP, CNNMAR-1, and CNNMAR-2 reference images with exemplary ROIs shown in red boxes. (b) Standard deviation within the ROIs was computed for both CNNMAR methods. Higher values indicate an increased presence of artifact.

4.2 *Results on Sparse Data with Simulated Metal*

Figure 10 shows the performance of the CNNMAR methods on metal data simulated using the polyenergetic simulation method shown in Fig. 5. Both methods remove nearly all metal artifacts from the image, yielding a highly accurate reconstruction of the screws; however, due to the sparse view acquisition strategy, reduced contrast in soft-tissue and bony structures is observed for both methods. The boxplots of Fig. 10(b) show the RMSE between the CNNMAR outputs and the ground truth images in ROIs near the screw locations (with metal and air regions removed from the RMSE computation). A small but significant ($p < 0.005$, paired t-test) reduction in RMSE is observed using the CNNMAR-2 method, again indicating that it provides greater preservation of anatomical structure ($3.7 \times 10^{-3} \pm 0.4 \times 10^{-3}$ mm^{-1} (mean ± std) for CNNMAR–1 vs. $3.4 \times 10^{-3} \pm 0.4 \times 10^{-3}$ mm^{-1} for CNNMAR-2).

Figure 11 illustrates the presence of metal blooming for both CNNMAR-1 and CNNMAR-2 in the simulated metal dataset. Both methods markedly reduce the metal blooming, clearly revealing details in the screw threads. The boxplots in Fig. 11(b) shows a small but significant ($p < 0.005$, paired t-test) reduction in the blooming ratio using CNNMAR-2 (1.15 mean blooming ratio) versus CNNMAR-1 (1.19 mean blooming ratio).

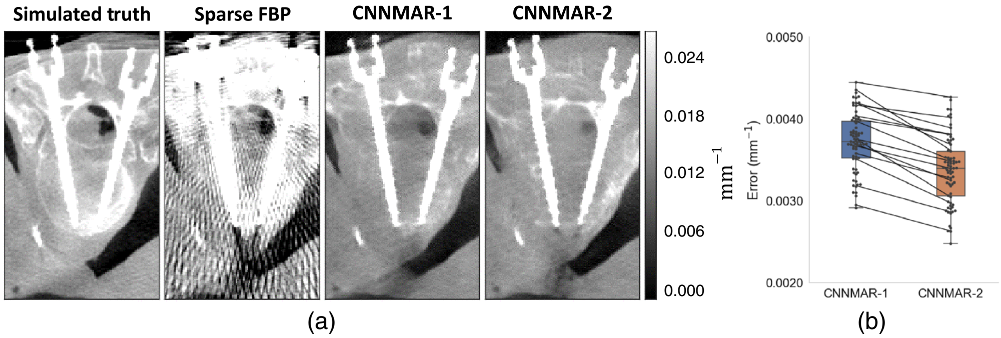

Fig. 10 Performance of CNNMAR methods in data with simulated metal screws. (a) From left-to-right, exemplary images of the monoenergetic fully sampled ground truth, polyenergetic Sparse FBP, CNNMAR-1 output, CNNMAR-2 output. (b) Boxplots depicting the RMSE error of the CNNMAR methods compared to ground truth.

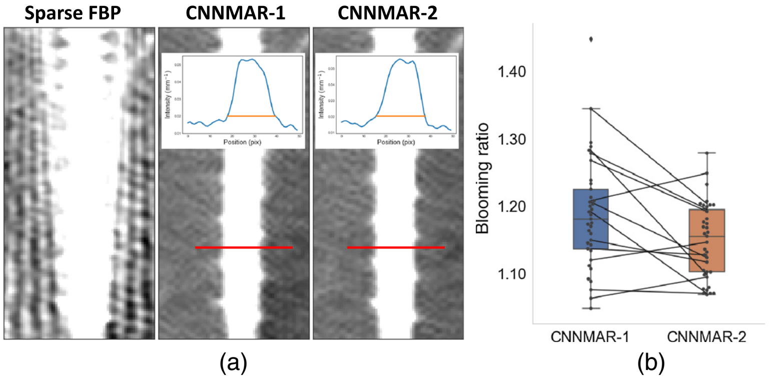

Fig. 11 Blooming ratio evaluation of CNNMAR methods applied on sparse scans with simulated polyenergetic screws. (a) Exemplary images with Sparse FBP reference and CNNMAR-1 and CNNMAR-2 outputs. Red lines depict an exemplary cross-section for evaluation the screw diameter, with the inset plots showing line profiles with the diameter computed using the full width at 20% of the maximum. (b) Boxplots depict the distribution of the blooming ration among the dataset for both CNNMAR methods.

4.3 *Results on Sparse Data at Reduced Tube Current*

Reducing the tube current in sparse-view image acquisition allows further reduction of the total dose of the scan while maintaining the 2-deg angular sampling of the sparse-view acquisition. Figure 12(a) shows the performance of the CNNMAR-2 method as the mA is lowered (using simulated noise injection) from 25 to 1.56 mA, where the top and bottom row show nearby axial slices (with the top row containing pedicle screws). Compared to the 1.56 mA Sparse FBP, the 1.56 mA CNNMAR-2 output greatly reduces the image noise (from 6.1×10^{-3} to 1.6×10^{-3} mm^{-1} noise standard deviation) with only a slight decrease in the measured edge-spread-function width (σ_{esf}) from 0.5 to 0.6 mm, where a comparable reduction in noise using a Gaussian filter (Sparse FBP: Blur) would increase σ_{esf} to 0.9 mm. Therefore, the noise reduction associated with CNNMAR comes with only a minimal expense of spatial resolution. Furthermore, qualitative inspection of the 1.56-mA CNNMAR-2 output shows similar performance in artifact reduction compared to the 25-mA CNNMAR-2 reconstruction.

Figure 12(b) plots the noise (standard deviation in voxel values in an otherwise uniform region) as the tube current is lowered down to 1/32nd of the nominal mA for the scan. The noise for CNNMAR-2 remains under 2×10^{-3} mm^{-1} across all tube current levels, outperforming Sparse FBP at its highest dose level. Therefore, tube current reduction presents a promising means to further reduce dose in a manner compatible with CNNMAR-2 workflow (with benefits in metal as well as non-metal regions).

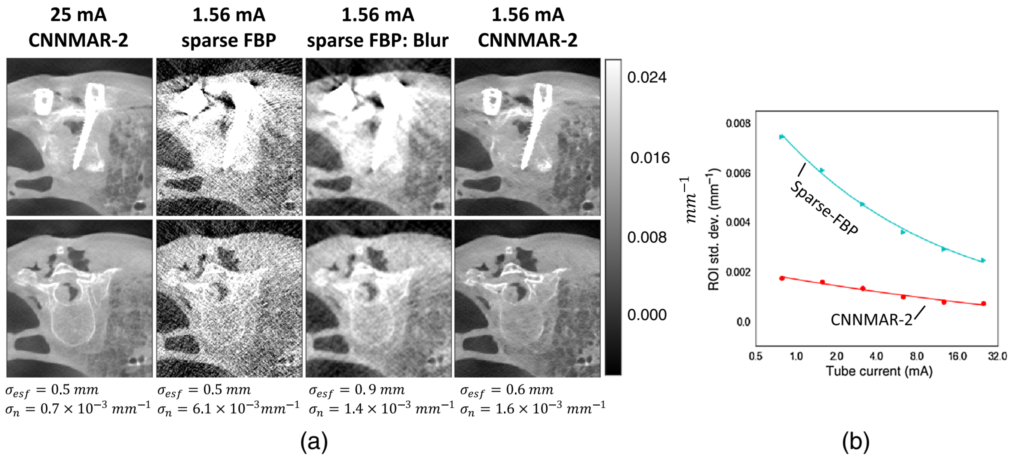

Fig. 12 Evaluation of CNNMAR-2 at reduced tube current. (a) Visualization of CNNMAR-2 performance at 1/16th the nominal tube current. Top and bottom rows depict two nearby axial slices. Columns from left to right show CNNMAR-2 at 25 mA, Sparse FBP at 1.56 mA, Sparse FBP at 1.56 mA with Gaussian blur, and CNNMAR-2 and 1.56 mA. Below each column show the measured ESF width noise standard deviation. (b) Plot of the tube current vs. noise standard deviation for CNNMAR-2 and Sparse FBP.

5 Conclusion

In this work, we proposed and validated a dual CNN approach for learning-based MAR that is physically motivated to address factors of data sparsity and polyenergetic beam-hardening and designed to perform on the large 3D datasets associated with CBCT acquisitions. While both CNNMAR-1 and CNNMAR-2 performed well at reducing metal and sparse-view artifacts, with CNNMAR–2 performing better in regions of particularly severe streaking, the use of the two-domain approach with CNNMAR-2 better preserved the underlying anatomy. Due to the particularly severe artifacts in regions between two pedicle screws, CNNMAR-1 was observed (Figs. 6 and 8) to often remove artifact without recovering the underlying anatomy, leaving regions with blurred out. The sinogram-domain network of CNNMAR-2, however, was shown to provide sufficient beam-hardening compensation prior to reconstruction (Fig. 6), such that underlying anatomy was better recoverable by the subsequent image-domain network. It is worth emphasizing that while the training data (containing only randomized screw positions) did not explicitly model this two-screw scenario, the CNNMAR-2 network was able to generalize well to this challenging setting in the test data.

The total runtime for reconstructing the $512 \times 512 \times 192$ voxel volume with CNNMAR-2 on an Nvidia Titan RTX GPU is 27.8 s: divided among 12.5 s for the sinogram-domain network, 7.8 s for FBP, and 7.5 s for the image-domain network. On the other hand, CNNMAR-1 required only 14.9 s for reconstruction (7.4 s for FBP and 7.5 s for the image-domain network). The CNNMAR–1 method has the further benefit of not requiring the raw projection data (only the reconstructed data). As CNNMAR-1 provides comparable performance to CNNMAR-2, the benefits in runtime and input data may make it better suited for some applications at some expense of resolvable anatomic structure.

A novel contribution of this work, outside of using a learning-based method to address both metal and view-sampling artifacts, was to design and realize a method for correction of strong metal artifacts in sparse-view data through the linearization of metal regions in the sinogram. While most sinogram-based MAR techniques use a metal segmentation to completely remove the metal (and add a visualization of the metal after reconstruction), the technique of CNNMAR-2 directly linearizes the metal attenuation in the sinogram (while also accounting for the small scatter term) [described by Eqs. (4) and (5)] so that the metal is naturally reconstructed along with the surrounding anatomy. This permits a non-iterative approach that does not rely on segmenting the metal and absolves the challenge of iterating from an initial image that is strongly degraded by truncation and intense metal and view-sampling artifacts.

All of the cadaver studies in this work involved abdominal and thoracic scans of adult specimens. With a 20-cm FOV for the CBCT system, the full lateral extent of such anatomy was truncated, so a simple model of the exterior boundary (32-cm diameter of water) was used to approximate the source energy spectrum for metal simulation for all projections. Such a value deserves modification for other anatomical sites (e.g., head or limbs) or pediatric cases. An interesting area of future work includes view-dependent estimation of the patient thickness from the projection data.

The work above specifically addressed metal artifacts arising from pedicle screws, with both the training and evaluation carried out using images containing various sizes of these screws. Extension of the training framework to other metal hardware is straight forward, requiring only 3D models and the material compositions of the metal components; however, the ability of the network to generalize to metal objects not observed during training deserves further investigation. As discussed in Sec. 3.3, the model of Eqs. (1) and (3) contains a residual bias in the regions where the shaft and tulip overlap in the projections. In the region of tulip-shaft overlap, the bias relates to a ~2.3% error in the projection intensity and a ~3.3% error in reconstructed metal intensity (relative to surrounding anatomy), with no visually discernable difference in the metal artifacts. Such an approximation will be worse for larger objects containing multiple metal materials.

While the work presented above addressed two of the primary sources of metal artifact, namely, sparse sampling and beam hardening, several other sources of artifact deserve further investigation, including incorporation of a physics-based x-ray scatter model within the training data.

Disclosures

The research was supported by the National Institutes of Health under Grant No. R01-EB-017226 and academic–industry collaboration with Medtronic.

References

1. B. De Man et al., “Metal streak artifacts in x-ray computed tomography: a simulation study,” *IEEE Trans. Nucl. Sci.* **46**(3), 691–696 (1999).
2. J. Johnson, A. Alahi, and L. Fei-Fei, “Perceptual losses for real-time style transfer and super-resolution,” in *Eur. Conf. Comput. Vision*, pp. 694–711 (2016).
3. Q. Yang et al, “Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss,” *IEEE Trans. Med. Imaging* **37**(6), 1348–1357 (2018).
4. L. A. Feldkamp, L. C. Davis, and J. W. Kress, “Practical cone-beam algorithm,” *J. Opt. Soc. Am.* **1**(6), 612–619 (1984).
5. W. A. Kalender, R. Hebel, and J. Ebersberger, “Reduction of CT artifacts caused by metallic implants,” *Radiology* **164**(2), 576–577 (1987).
6. E. Meyer et al., “Normalized metal artifact reduction (NMAR) in computed tomography,” *Med. Phys.* **37**(10), 5482–5493 (2010).
7. I. A. Elbakri and J. A. Fessler, “Segmentation-free statistical image reconstruction for polyenergetic x-ray computed tomography with experimental validation,” *Phys. Med. Biol.* **48**(15), 2453 (2003).
8. A. Uneri et al., “Known-component metal artifact reduction (KC-MAR) for cone-beam CT,” *Phys. Med. Biol.* **64**(16), 165021 (2019).
9. X. Zhang et al., “Known-component 3D image reconstruction for improved intraoperative imaging in spine surgery: a clinical pilot study,” *Med. Phys.* **46**, 3483–3495 (2019).
10. P. Wu et al., “C-arm orbits for metal artifact avoidance (MAA) in cone-beam CT,” *Phys. Med. Biol.* **65**, 165012 (2020).
11. G. J. Gang, J. H. Siewerdsen, and J. W. Stayman, “Non-circular CT orbit design for elimination of metal artifacts,” *Proc. SPIE* **11312**, 1131227 (2020).
12. G. H. Weiss, A. J. Talbert, and R. A. Brooks, “The use of phantom views to reduce CT streaks due to insufficient angular sampling,” *Phys. Med. Biol.* **27**, 1151–1162 (1982).

13. M. Bertram et al., "Directional view interpolation for compensation of sparse angular sampling in cone-beam CT," *IEEE Trans. Med. Imaging* **28**(7), 1011–1022 (2009).
14. S. Niu et al., "Sparse-view x-ray CT reconstruction via total generalized variation regularization," *Phys. Med. Biol.* **59**(12), 2997 (2014).
15. J. Bian et al., "Evaluation of sparse-view reconstruction from flat-panel-detector cone-beam CT," *Phys. Med. Biol.* **55**(22), 6575 (2010).
16. L. Gjesteby et al., "Deep neural network for CT metal artifact reduction with a perceptual loss function," in *Proc, Fifth Int. Conf. Image Formation in X-ray Comput. Tomogr.* (2018).
17. S. Xu and H. Dang, "Deep residual learning enabled metal artifact reduction in CT," *Proc. SPIE* **10573**, 105733O (2018).
18. M. U. Ghani and W. C. Karl, "Deep learning based sinogram correction for metal artifact reduction," *Electron. Imaging* **2018**(15), 4721–4728 (2018).
19. H. S. Park et al., "CT sinogram-consistency learning for metal-induced beam hardening correction," *Med. Phys.* **45**(12), 5376–5384 (2018).
20. H. Liao et al., "Generative mask pyramid network for CT/CBCT metal artifact reduction with joint projection-sinogram correction," *Lect Notes Comput. Sci.* **11769**, 77–85 (2019).
21. Y. Zhang and H. Yu, "Convolutional neural network based metal artifact reduction in x-ray computed tomography," *IEEE Trans. Med. Imaging* **37**(6), 1370–1381 (2018).
22. W.-A. Lin et al., "DuDoNet: dual domain network for CT metal artifact reduction," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 10512–10521 (2019).
23. Z. Zhang et al., "A sparse-view CT reconstruction method based on combination of DenseNet and deconvolution," *IEEE Trans. Med. Imaging* **37**(6), 1407–1417 (2018).
24. M. Abadi et al., "TensorFlow: a system for large-scale machine learning," in *Proc. 12th USENIX Conf. Oper. Syst. Design and Implement.*, USENIX Association, Berkeley, California, pp. 265–283 (2016).
25. D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent.* (2015).
26. G. Huang et al., "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 4700–4708 (2017).
27. O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," *Lect. Notes Comput. Sci.* **9351**, 234–241 (2015).
28. A. S. Wang et al., "Low-dose preview for patient-specific, task-specific technique selection in cone-beam CT," *Med. Phys.* **41**(7), 71915 (2014).
29. J. Punnoose et al., "spektr 3.0—a computational tool for x-ray spectrum modeling and analysis," *Med. Phys.* **43**(8Part1), 4711–4717 (2016).
30. J. Tan et al., "Sharpness preserved sinogram synthesis using convolutional neural network for sparse-view CT imaging," *Proc. SPIE* **10949**, 109490E (2019).
31. Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *37th Asilomar Conf. Signals, Syst. & Comput.*, Vol. 2, pp. 1398–1402 (2003).
32. A. Sisniega et al., "Motion compensation in extremity cone-beam CT using a penalized image sharpness criterion," *Phys. Med. Biol.* **62**(9), 3712 (2017).

Biographies of the other authors are not available.