

# Deep learning-based motion compensation for four-dimensional cone-beam computed tomography (4D-CBCT) reconstruction

Zhehao Zhang<sup>1</sup> | Jiaming Liu<sup>2</sup> | Deshan Yang<sup>3</sup> | Ulugbek S. Kamilov<sup>2,4</sup> | Geoffrey D. Hugo<sup>1,4</sup>

<sup>1</sup>Department of Radiation Oncology, Washington University School of Medicine in St. Louis, St. Louis, Missouri, USA

<sup>2</sup>Department of Electrical and Systems Engineering, Washington University in St. Louis, St. Louis, Missouri, USA

<sup>3</sup>Department of Radiation Oncology, Duke University School of Medicine, Durham, North Carolina, USA

<sup>4</sup>Department of Computer Science and Engineering, Washington University in St. Louis, St. Louis, Missouri, USA

## Correspondence

Geoffrey D. Hugo, Department of Radiation Oncology, Washington University School of Medicine in St. Louis, 4921 Parkview Place, St. Louis, MO 63110, USA.  
Email: [gdhugo@wustl.edu](mailto:gdhugo@wustl.edu)

## Funding information

National Institute of Biomedical Imaging and Bioengineering, Grant/Award Number: R03-EB028427; National Heart, Lung, and Blood Institute, Grant/Award Number: R01-HL148210

## Abstract

**Background:** Motion-compensated (MoCo) reconstruction shows great promise in improving four-dimensional cone-beam computed tomography (4D-CBCT) image quality. MoCo reconstruction for a 4D-CBCT could be more accurate using motion information at the CBCT imaging time than that obtained from previous 4D-CT scans. However, such data-driven approaches are hampered by the quality of initial 4D-CBCT images used for motion modeling.

**Purpose:** This study aims to develop a deep-learning method to generate high-quality motion models for MoCo reconstruction to improve the quality of final 4D-CBCT images.

**Methods:** A 3D artifact-reduction convolutional neural network (CNN) was proposed to improve conventional phase-correlated Feldkamp–Davis–Kress (PCF) reconstructions by reducing undersampling-induced streaking artifacts while maintaining motion information. The CNN-generated artifact-mitigated 4D-CBCT images (CNN enhanced) were then used to build a motion model which was used by MoCo reconstruction (CNN+MoCo). The proposed procedure was evaluated using in-vivo patient datasets, an extended cardiac-torso (XCAT) phantom, and the public SPARE challenge datasets. The quality of reconstructed images for XCAT phantom and SPARE datasets was quantitatively assessed using root-mean-square-error (RMSE) and normalized cross-correlation (NCC).

**Results:** The trained CNN effectively reduced the streaking artifacts of PCF CBCT images for all datasets. More detailed structures can be recovered using the proposed CNN+MoCo reconstruction procedure. XCAT phantom experiments showed that the accuracy of estimated motion model using CNN enhanced images was greatly improved over PCF. CNN+MoCo showed lower RMSE and higher NCC compared to PCF, CNN enhanced and conventional MoCo. For the SPARE datasets, the average ( $\pm$  standard deviation) RMSE in  $\text{mm}^{-1}$  for body region of PCF, CNN enhanced, conventional MoCo and CNN+MoCo were  $0.0040 \pm 0.0009$ ,  $0.0029 \pm 0.0002$ ,  $0.0024 \pm 0.0003$  and  $0.0021 \pm 0.0003$ . Corresponding NCC were  $0.84 \pm 0.05$ ,  $0.91 \pm 0.05$ ,  $0.91 \pm 0.05$  and  $0.93 \pm 0.04$ .

**Conclusions:** CNN-based artifact reduction can substantially reduce the artifacts in the initial 4D-CBCT images. The improved images could be used to enhance the motion modeling and ultimately improve the quality of the final 4D-CBCT images reconstructed using MoCo.

## KEYWORDS

4D-CBCT, deep learning, motion compensation

## 1 | INTRODUCTION

Cone-beam computed tomography (CBCT) imaging systems have been widely used in image-guided radiation therapy (IGRT) to ensure correct patient setup and increase the precision of delivering radiation dose to the target.<sup>1</sup> For linear accelerator (LINAC) gantry mounted CBCT, densely sampled projections within a 180–360° gantry rotation are required for CBCT imaging.<sup>2</sup> Due to the limited rotation speed of a conventional LINAC gantry, the acquisition time of CBCTs can be 60 s or even longer, allowing collection of data over many respiratory cycles of a patient. For those body anatomical regions susceptible to respiratory motion, like thorax and abdomen, inconsistent locations of anatomy in the projections will result in motion blurring and artifacts, and cause distortions of the tumor target volumes and incorrect positional information.<sup>3</sup>

To improve the accuracy of image-guidance in the presence of breathing motion, four-dimensional (4D) CBCT has been developed to provide time-resolved target localization.<sup>4</sup> All cone-beam projections are first retrospectively sorted into different phase bins according to their position in the respiratory cycle when acquired, then multiple phase-correlated 3D-CBCT images are reconstructed with the phase-binned subsets of projections. Although 4D-CBCT has shown the capability of reducing the motion artifacts, phase-correlated CBCTs reconstructed using conventional Feldkamp–Davis–Kress (FDK) algorithm,<sup>5</sup> the clinical standard CBCT reconstruction method, suffer from severe streaking artifacts and significantly increased image noise due to insufficient projection images and large angular spacings between projections resulted by the phase binning process.<sup>6</sup>

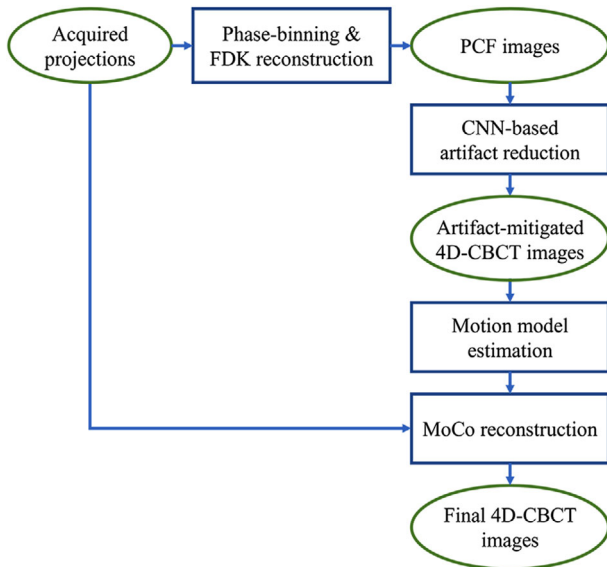
Many efforts have been devoted to addressing these limitations. One approach to relieving the degraded image quality is to adopt slow gantry rotation or to use multiple gantry rotations.<sup>7,8</sup> This approach will lead to much longer acquisition time (3–4 min) and more imaging dose to the patient. An alternative way to improve image quality is using iterative reconstruction algorithms with regularization terms, for example, total variation (TV) minimization.<sup>9</sup> Nevertheless, iterative reconstruction often leads to over-smoothness and losses of fine structures. Methods have been proposed to enhance the reconstruction by incorporating prior knowledge, such as the prior image constraint compressive sensing (PICCS),<sup>10</sup> where another regularization term is enforced by minimizing the TV of the difference between the phase-correlated image and a prior image reconstructed using all available projections. Since PICCS is only a trade-off, residual motion has still been observed in the reconstructed 4D-CBCT.<sup>11</sup> 4D reconstruction using spatial and temporal regularization (ROOSTER)<sup>12</sup> has shown strong performance by enforcing the consistency between different phases and reconstructing all

phase-correlated images at once, where more information is exploited for the reconstruction of each phase. However, like other iterative algorithms, ROOSTER is still sensitive to the parameters selection and suffers from high computational demands.

Motion-compensated (MoCo) reconstruction is another strategy, which attempts to compensate for the respiratory motion of each phase-correlated images by utilizing all available projections with the help of inter-phase deformation vector fields (DVF) via a motion model. Earlier MoCo methods estimate the DVFs from the 4D planning CT.<sup>13,14</sup> However, the prior motion model built from the 4D planning CT may not remain valid over the treatment course due to variations of anatomy and/or the motion pattern. Many other methods have also been proposed to directly estimate the motion model from on-board 4D-CBCT images.<sup>15–17</sup> However, their performance has been limited by the inaccuracy of motion models generated from low-quality 4D-CBCT images. Improving the initial 4D-CBCT reconstructions may enable generation of high-quality motion models for use in later MoCo reconstructions.

Over the last few years, deep learning has been widely used in medical image reconstruction.<sup>18</sup> As for the field of 4D-CBCT, Zhi et al.<sup>19</sup> incorporated a prior image reconstructed from all projections into a convolutional neural network (CNN) to enhance phase-correlated CBCT images. Jiang et al.<sup>20</sup> proposed using a symmetric residual CNN to learn a restoring pattern from under-sampled images reconstructed using TV regularization to the ground truth. Madesta et al.<sup>21</sup> trained a CNN between pseudo-average and time-average CBCT images to boost the image quality of 4D-CBCT. These methods started with the iterative reconstruction results to learn the relationships between under-sampled images and corresponding high-quality images or ground truth, which relied on the availability of such high-quality ground truth images and could not recover the details already distorted or lost in the under-sampled images. Huang et al.<sup>22</sup> utilized CNN-based models to fine-tune the DVF for MoCo and improve the efficiency and accuracy of 4D-CBCT reconstruction. However, the DVF estimation based on low-quality initial images was still required.

In this study, we combined the deep learning and MoCo reconstruction from a new perspective to address the 4D-CBCT reconstruction problem. We hypothesized that high-quality initial 4D-CBCTs would facilitate the accurate estimation of motion model which enables a MoCo reconstruction resulting in improved image quality. Instead of using deep learning directly as a post-processing operator or to fine-tune the estimated DVFs, we proposed a novel 4D-CBCT reconstruction procedure that enhances the MoCo reconstruction by applying deep learning-based artifact reduction to improve the quality of initial images. The proposed procedure consists of three steps: (1) to use a 3D CNN to



**FIGURE 1** Flowchart of the proposed 4D-CBCT reconstruction procedure

reduce the structural artifacts (e.g., streaks) from initial FDK-reconstructed 4D-CBCTs, (2) to generate a high-quality motion model using the artifact-mitigated initial images, and (3) to use the high-quality motion model and MoCo to re-reconstruct the final 4D-CBCTs, which may further restore the information lost during under-sampled FDK reconstruction. This procedure eliminates the need for a pre-treatment motion model by utilizing a data-driven posteriori motion modeling. For evaluation of the proposed procedure, 4D-CBCT reconstructions of real patient data, phantom data and the publicly available data from the SPARE-view REconstruction (SPARE) Challenge for 4D-CBCT<sup>23</sup> were carried out.

## 2 | METHODS

### 2.1 | Proposed procedure

#### 2.1.1 | Overview

The overall flowchart of our proposed procedure is presented in Figure 1. First, all CBCT projections acquired from a single free-breathing scan were binned into 10 different breathing phases based on a respiratory signal. Then, individual reconstruction for each phase was performed using a conventional phase-correlated FDK (PCF) algorithm, which only back-projected those projections corresponding to the given phase bin. A pre-trained CNN was then used to reduce structural artifacts to improve the image quality of the PCF reconstructions. Then, a set of DVFs was estimated based on these artifact-mitigated 4D-CBCT images via deformable registrations which was subsequently used as the motion model. In the final step, MoCo reconstruction was per-

formed to reconstruct each phase of 4D-CBCT with the motion model and all available projections.

#### 2.1.2 | Projection binning and initial FDK reconstruction

For one 4D-CBCT scan, let  $P = \{p_1, p_2, \dots, p_{N_p}\} \in \mathbb{R}^3$  represent the total  $N_p$  air-normalized and log-transformed projections acquired along a circular axial orbit around the patient. Since FDK was used as the standard reconstruction method in this study, without considering patient motion during the acquisition, the 3D reconstruction can be described as:

$$x = F^{-1}P, \quad (1)$$

where  $F$  represents the x-ray forward projection operator,  $F^{-1}$  is FDK reconstruction operator, and  $x$  is the reconstructed free-breathing 3D image which should be contaminated by respiratory motion. In this study, all FDK-based reconstruction algorithms were implemented using the open source Reconstruction Toolkit (RTK, v2.1.0).<sup>24</sup>

Assuming the respiratory cycle can be divided into  $n$  breathing phases, to achieve the phase binning for the following 4D reconstruction, each 2D projection from  $P$  will be assigned into one of  $n$  phase bins based on the respiratory signal, labeled as  $P_i$  with  $i = 1, 2, \dots, n$ . The PCF reconstruction of image  $x_i$ , or linear attenuation coefficients of 4D-CBCT at phase  $i$ , is given by:

$$x_i = F^{-1}P_i, \quad (2)$$

The PCF images can be regarded as containing minimal respiratory motion, but they will also suffer from severe streaking artifacts and increased image noise due to the undersampling and large angular spacings of projections normally present when clinical CBCT scans are used for 4D-CBCT reconstruction.

#### 2.1.3 | Self-contained deep learning-based 3D artifact reduction

To improve the quality of initial 4D-CBCT images, a CNN was trained to learn the relationship between low-quality, streaking artifact-corrupted 3D CBCT images and high-quality 3D CBCT images with the same patient position and respiratory state.

To overcome the lack of ground truth 4D-CBCT images, we constructed the training dataset using a self-contained way inspired by Madesta et al.<sup>21</sup> Unlike the phase binning for 4D-CBCT, a specific projection selection scheme was applied to sort 2D projections corresponding to different breathing phases under different respiratory cycles into one of  $n$  different

pseudo-average projection subsets, denoted by  $P_i^{pa}$  with  $i = 1, 2, \dots, n$ . Specifically, for example, subset  $P_{i=1}^{pa}$  contains projections corresponding to the first breathing phase in the first respiratory cycle, the second breathing phase in the second respiratory cycle, and so on. While projections belonging to  $P_{i=2}^{pa}$  come from the second breathing phase in the first respiratory cycle, the third breathing phase in the second respiratory cycle, and so on. Following this procedure, pseudo-average CBCTs  $x_i^{pa}$  were reconstructed as:

$$x_i^{pa} = F^{-1}P_i^{pa}, \quad (3)$$

The artifact characteristics and image quality of pseudo-average image  $x_i^{pa}$  will be similar to those of the PCF image  $x_i$ . In addition,  $x_i^{pa}$  also features the same averaged respiratory motion as the 3D image  $x$  does, since utilized projections come from all breathing phases. Therefore, the network could be trained to reduce the streaking artifacts by regarding  $(x_i^{pa}, x)$  as the training pair, where pseudo-average  $x_i^{pa}$  served as streaking-corrupted image with respiratory motion and the 3D time-average  $x$  served as the corresponding high-quality streaking-free representation with the same motion.

As CBCT is an intrinsically volumetric imaging method, a 3D network which can leverage inter-slice spatial information was utilized in this study. Since directly processing the entire volume can be intractable because of graphics processing unit (GPU) memory limitation, we adopted patch-wise training by extracting overlapped 3D patches from the training volume pairs. It was notable that such a patch-based strategy has also shown its advantages in detecting perceptual differences of local regions and mitigating over-fitting by significantly increasing the number of samples.<sup>25</sup> Due to the initial 3D image quality varying with slice position,<sup>26</sup> our extracted patches were design to arrange along the transverse plane and always cover all slices.

After generating a set of paired 3D patches  $\{(x_i^{pa}|_k^j, x|_k^j)\}$ , the artifact-reduction CNN parameterized by  $\theta$  can be trained through minimizing the  $\ell_2$  loss over the training set:

$$\operatorname{argmin}_{\theta} \sum_{k,i,j} \left\| \operatorname{CNN}_{\theta} \left( x_i^{pa}|_k^j \right) - x|_k^j \right\|_2^2, \quad (4)$$

where  $k$  denotes the index of the CBCT scan utilized for training,  $i$  denotes the index of pseudo-average image for one given scan and  $j$  denotes the index of extracted patch pair. Once the CNN was trained to transform low-quality into high-quality average CBCTs, it was used to improve initial PCF images by reducing the streaking artifacts. The output of the network

is an artifact-mitigated prediction ( $\hat{x}_i$ ) of each  $x_i$  for each phase  $i$ . The general training and testing procedures of the artifact-reduction CNN is shown in Figure 2.

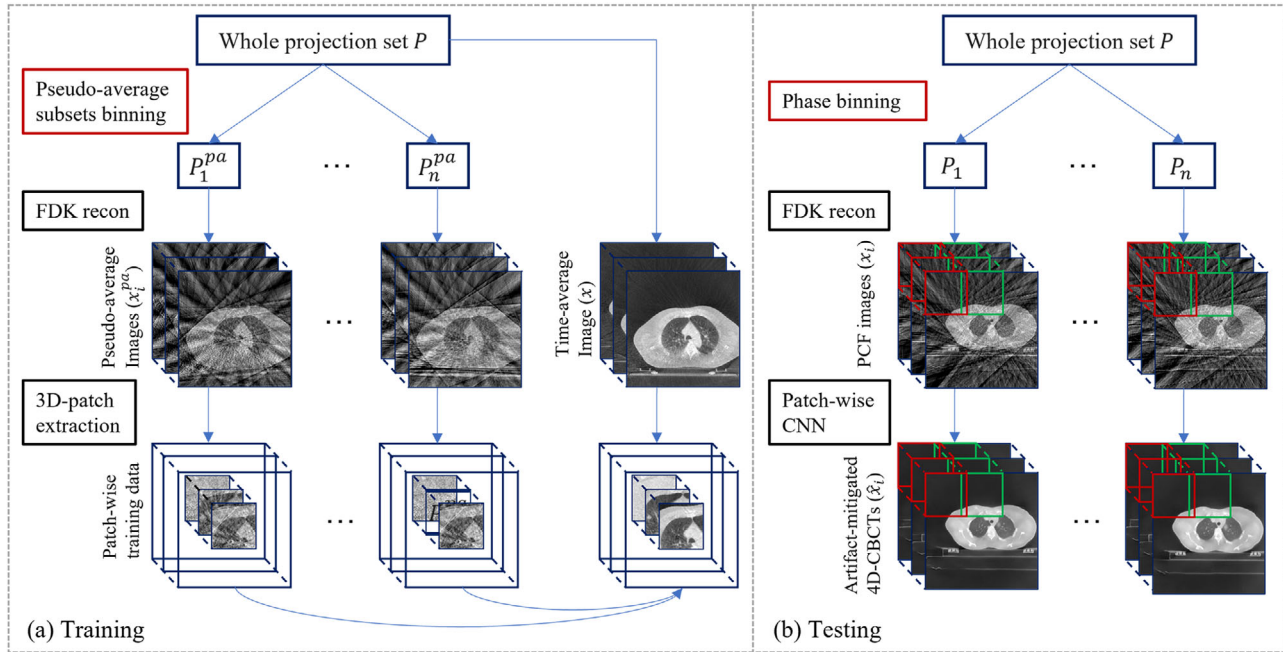
#### 2.1.4 | Network architecture

A 3D U-Net<sup>27</sup> with modifications was used for implementing this artifact reduction pipeline. Figure 3 illustrates the proposed network architecture. 3D convolutions with the reflection padding were utilized in both contracting and expansive stages, each followed by a batch normalization (BN) and a rectified linear unit (ReLU). To avoid the potential checkerboard artifacts introduced by transposed convolution,<sup>28</sup> upsampling was achieved by nearest-neighbor interpolation followed by a  $2 \times 2 \times 2$  convolution in the expansive path. In addition, residual learning was applied through a residual connection to subtract the input from the output of the final layer, based on the assumption that the residual noise and artifact components between inputs and targets are much easier to be fitted than directly predicting the latent artifact-free images.<sup>29</sup> The details of network training and testing are described below in the Experiments setup subsection.

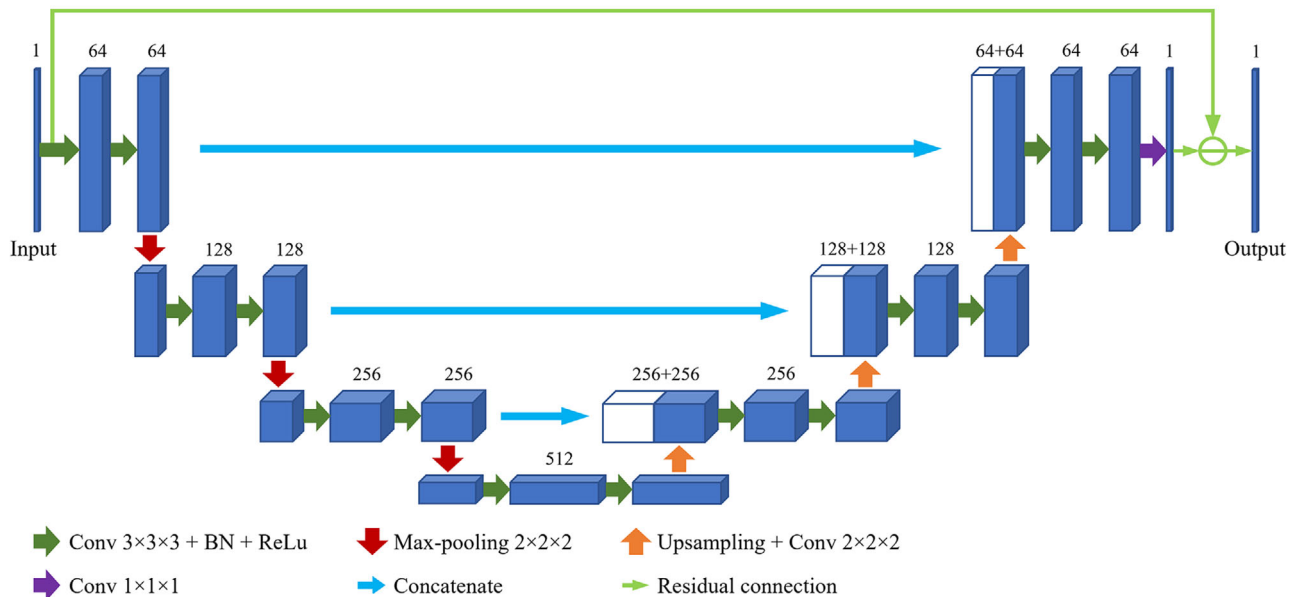
#### 2.1.5 | Motion estimation and motion-compensated reconstruction

As we have mentioned previously, the quality of initial PCF images is severely degraded due to insufficient number of projections in each breathing phase bin. Some small structures can be totally overwhelmed by the noise and streaking artifacts. Although most artifacts can be removed from PCFs by the trained CNN, it would be hard or even impossible to recover those tiny structures only using CNN as a post-processing operator for artifact reduction. However, due to the high spatiotemporal correlation of 4D-CBCTs, the motion will present low frequency and smoothness such that high-spatial-frequency tiny structures will not be essential for the motion estimation.

To ensure the data fidelity and restore the information which has been impaired during PCF, MoCo reconstruction was performed using the artifact-mitigated 4D-CBCT images ( $\hat{x}_i$ ) from the CNN to build the motion model. As shown in the overall workflow of our proposed method in Figure 1, a groupwise 4D registration with fixed-reference frame approach<sup>17</sup> was first utilized to estimate the DVFs. All registrations were implemented using elastix v5.0.1<sup>30</sup> with a hierarchical registration scheme and cubic B-spline transformations. Automatically-generated anatomical masks were used to confine the registration within the body region to



**FIGURE 2** Workflow of the artifact-reduction CNN training (a) and testing (b) procedures. In training procedure, whole projection set is divided into  $n$  pseudo-average subsets, each containing projections corresponding to all breathing phases.  $n$  pseudo-average images ( $x_i^{pa}$ ) reconstructed from each subset paired with the time-average image ( $x$ ) reconstructed from the whole projection set constitute the training data. Patch-wise training is utilized, where each 3D-patch covers partial transverse plane but all slices. While in testing phase, whole projection set is separated into  $n$  phase bins, each only containing projections corresponding to one breathing phase.  $n$  PCF images ( $x_i$ ) reconstructed from each subset are used as input for the trained patch-wise CNN to get artifact-mitigated 4D-CBCT images ( $\hat{x}_i$ )



**FIGURE 3** Architecture of applied 3D U-Net with residual learning for artifact reduction. The cubic blocks represent feature maps with denoted number of channels

eliminate the influence of background artifacts. Then, the MoCo reconstruction for each breathing phase was performed using a voxel-driven FDK algorithm which back-projected the projection data into the volume space along warped trajectories according to the phase-

correlated DVFs.<sup>14,31</sup> Note that different from the PCF, all acquired projections will be used for back-projection in MoCo reconstruction for each phase. Our code and model parameters are available at <https://github.com/zhangzhehao95/CNN-MoCo>.

## 2.2 | Datasets

To evaluate the performance of the proposed procedure, three types of different datasets were utilized—in-vivo patient datasets of 18 patients which contain the training data for our proposed artifact-reduction CNN, a digital phantom dataset, and the Varian subset of the SPARE challenge dataset.

### 2.2.1 | In-vivo patient dataset

With Institutional Review Board approval, a dataset of free-breathing CBCTs was collected retrospectively for 18 patients undergoing radiation therapy for lung and upper abdomen tumors. For each patient, data for three to five clinical CBCT scans were collected after the radiotherapy treatment sessions on a Varian Edge system (Varian Medical Systems, Palo Alto, California, USA). Patients were scanned for image guidance using a clinical standard thorax CBCT protocol, 60 s scan time, image acquisition frame rate about 15 frames/s. Around 900 projections per scan with a matrix size of  $1024 \times 768$  and a pixel spacing of  $0.338 \times 0.338 \text{ mm}^2$  were acquired using half bow-tie filter in a single  $360^\circ$  gantry arc. Source-to-isocenter distance was 1000 mm, source-to-detector distance was 1500 mm, and the detector was laterally shifted 160 mm for half-fan mode acquisition to increase the size of the field of measurement. The respiratory signal was directly extracted from the acquired projection data using Amsterdam Shroud method<sup>32</sup> with the RTK implementation.

The in-vivo patient dataset was randomly separated into training set (16 patients) and testing set (two patients). The training set was used for model training and hyperparameter tuning. Three CBCT scans per patient were kept for subsequent experiments to avoid unbalanced data. Among the training set, the average ( $\pm$  standard deviation) number of respiratory cycles or projection clusters within one phase bin was  $26.6 \pm 9.6$ , which is more critical to the 4D reconstruction than the absolute number of projections.<sup>33</sup> 3D FDK and 4D pseudo-average reconstructions were performed on training set, while only 4D PCF reconstruction was applied to the testing set. In this study, the respiratory cycle was divided into 10 phases ( $n = 10$ ), and all images were reconstructed using a 2 mm isotropic voxel spacing with  $224 \times 224 \times 96$  volume size.

### 2.2.2 | Phantom dataset

Since no ground truth for 4D-CBCT reconstructions were available for in-vivo patient data, the 4D extended cardiac-torso (XCAT) phantom<sup>34</sup> which can provide realistic human anatomy was utilized for quantitative analysis. With the maximum diaphragm motion set

to 20 mm and the maximum chest anterior-posterior expansion set to 5 mm, 10 phase-correlated phantom volumes were obtained, mimicking one respiratory cycle. The volume size for each digital phantom was  $512 \times 512 \times 220$  with voxel size of  $1 \times 1 \times 1 \text{ mm}^3$ . To simulate a 1-min CBCT scan, 900 evenly distributed projections over  $360^\circ$  were generated from 10 volumes using Joseph's ray-driven algorithm<sup>35</sup> with the same geometry setting as the in-vivo patient data acquisition, coupled to a simulated respiratory signal with a rate of 19 respirations per minute. The noisy signal  $s_i$  at each detector bin  $i$  was simulated based on the following noise model<sup>36</sup>:

$$s_i = \text{Poisson}(I_i \exp(-p_i)) + \text{Normal}(0, \sigma_e^2), \quad (5)$$

where  $I_i$  is the incident x-ray intensity,  $p_i$  is the noise-free attenuation line integral and  $\sigma_e^2$  is the background electronic noise variance which was chosen as 10. The generated noisy projections were used for subsequent 4D-CBCT reconstruction with initial digital phantoms as ground truth.

### 2.2.3 | SPARE challenge dataset

To further evaluate our proposed method, five clinical Varian datasets from the SPARE challenge were also used. Each SPARE dataset consists of a fully sampled projections set with 2400–3600 projections and a downsampled projection sets with 680 projections along with their corresponding respiratory signals. Planning target volume (PTV) mask was also provided to mark the tumor region. The downsampled projections were used for the 4D reconstruction here, with the reference images reconstructed using the fully sampled projections to enable evaluation against a gold standard.

## 2.3 | Experiments setup

### 2.3.1 | Data pre-processing and artifact-reduction network training

Only the in-vivo patient dataset was used for training our artifact-reduction CNN. In pre-processing, 3D images were normalized by clipping to the 0.5 and 99.5 percentiles of the image intensity distribution. 3D patches of size  $64 \times 64 \times 96$  with an in-plane overlap of 6 were extracted from the normalized volumes, resulting in 7680 (16 patients  $\times$  3 treatment sessions  $\times$  10 breathing phases  $\times$  16 patches) training patch pairs. No extra data augmentation was performed in this study.

The CNN kernel weights were initialized using the Glorot uniform technique.<sup>37</sup> Adam optimizer<sup>38</sup> was used to minimize the loss function with the minibatch size of 5. The learning rate was initially set to  $1 \times 10^{-3}$  and multiplied with 0.9 every five epochs. 100 epochs

were utilized for network training. Our network was implemented using Keras with Tensorflow v2.0.0 as a backend and trained on a single NVIDIA Tesla V100S GPU.

### 2.3.2 | Evaluation

The testing datasets from the in-vivo patient datasets, XCAT phantom dataset as well as the SPARE datasets were used to evaluate our proposed method. For each dataset, conventional PCF reconstruction was first performed to provide a starting point for our evaluations. The pre-trained CNN was then applied on the PCF images to generate artifact-mitigated 4D-CBCT images, referred to as CNN enhanced images, where the impact of CNN-based artifact reduction can be assessed. To test the potential advantages of artifact-mitigated 4D-CBCT images in estimating DVFs as well as the potential contribution of MoCo reconstruction towards restoring high contrast resolution, MoCo reconstruction was separately performed based on initial PCF images (conventional MoCo) and CNN enhanced images (CNN+MoCo). To validate the ability of our proposed method fairly, the state-of-the-art ROOSTER reconstruction was also performed for the patient datasets.

For XCAT phantom dataset, the DVFs estimated from the phase-correlated digital phantoms can be regarded as ground truth for motion model. Comparing with the ground truth, we qualitatively evaluated the accuracy of DVFs estimated by the initial PCF images and the CNN enhanced images.

To quantitatively evaluate the quality of reconstructed image using different methods as well as comparing with other previous works which also utilized the SPARE dataset for evaluation, the root-mean-square-error (RMSE) and normalized cross-correlation (NCC) were used for the two datasets with a ground truth or gold standard, XCAT phantom and SPARE dataset, which are defined as follows:

$$\text{RMSE}(X, \hat{X}) = \sqrt{\frac{1}{M} \sum_{m=1}^M (X_m - \hat{X}_m)^2}, \quad (6)$$

where  $X$  denotes the reconstructed image,  $\hat{X}$  denotes the corresponding reference image or ground truth,  $m$  is the voxel index and  $M$  is the total voxel number of the volumetric image.

$$\text{NCC}(X, \hat{X}) = \frac{1}{M} \sum_{m=1}^M \frac{(X_m - \mu_X)(\hat{X}_m - \mu_{\hat{X}})}{\sigma_X \sigma_{\hat{X}}}, \quad (7)$$

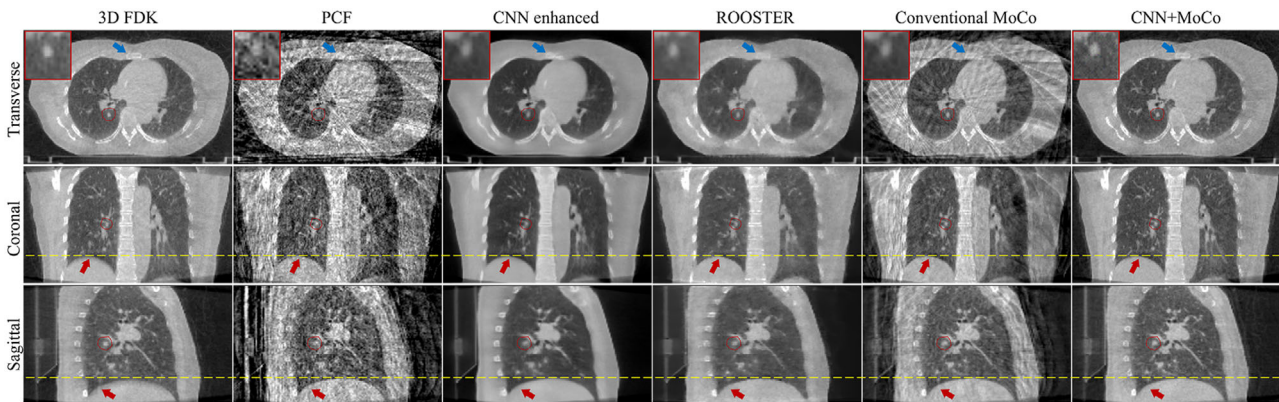
where  $\mu$  and  $\sigma$  represent the mean value and standard deviation of image.

## 3 | RESULTS

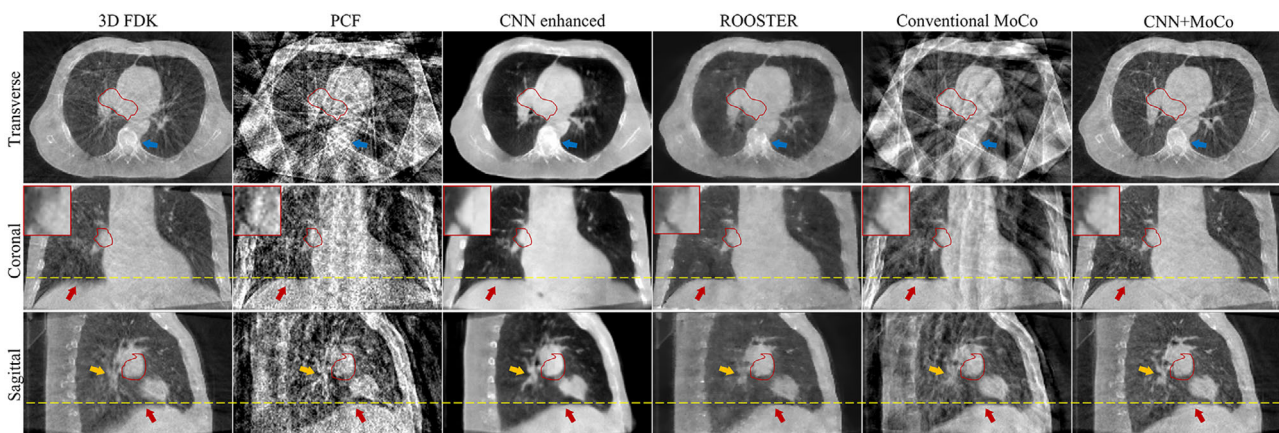
### 3.1 | In-vivo experiments

Figure 4 shows the reconstruction results of 3D FDK which utilized all available projections without binning, and PCF, CNN enhanced, ROOSTER, conventional MoCo as well as CNN+MoCo at the end-exhale phase of patient 1 from our in-vivo testing set. The tumor region is marked by red contours. The 3D FDK images were degraded by respiratory motion and regions affected by motion blurring are indicated by red arrows (diaphragm). The PCF images suffered from increased noise and severe streaking artifacts, while the pre-trained CNN can effectively mitigate the noise and streaking artifacts in the initial PCF reconstruction results. Nevertheless, slight distortion (see blue arrows) and over-smoothness (see red contours and zoomed-in images) can be observed. Iteration-based ROOSTER provided comparable image quality with CNN enhanced images, where over-smoothness can also be observed (blue arrows and red contours). It is interesting to note that ROOSTER took more than 10 min to finish the 4D reconstruction in our setting (with a GPU) and CNN enhanced only took about 1 min. Comparing PCF with conventional MoCo, although directly using the PCF images for motion compensation can improve the image quality, blurring and streaking artifacts remained and some fine structures disappeared due to the low quality of PCF images. CNN+MoCo, which compensates motion during image reconstruction using the motion information estimated on the artifact-mitigated 4D-CBCT images by CNN, can further enhance the image quality and recover the distorted or blurred regions.

Comparing to patient 1 who breathed for 36 cycles during the 1-min scan, patient 2 breathed much slower and had only 16 respiratory cycles during the scan. The motion-corrupted 3D FDK image and end-exhale phase images reconstructed with different methods of patient 2 are shown in Figure 5. Compared to patient 1, the blurring artifacts in the region of diaphragm (indicated by red arrows) and tumor (indicated by red contours and zoomed-in images) for 3D FDK were more obvious due to a larger range of respiratory motion. Additionally, the initial PCF images were also more degraded by streaking artifacts due to the larger angular separation of projection clusters. For this more challenging case, motion blurring artifacts were mitigated in the 4D-CBCT reconstructions. CNN can still successfully suppress the streaking artifacts of PCF, while conventional MoCo was unable to generate a visually acceptable result. ROOSTER also reduced the noise and streaking artifacts in the reconstructed images, but the lung vessels (yellow arrows) and spine region (blue arrows) were blurred due to the over-smoothness. The final CNN+MoCo reconstruction restored more fine structures compared to CNN enhanced images



**FIGURE 4** Reconstruction results of end-exhale phase of patient 1 from in-vivo testing set, using phase-correlated FDK (PCF, column 2), artifact-reduction CNN with PCF as input (CNN enhanced, column 3), ROOSTER (column 4), MoCo reconstruction with a motion model generated from PCF images (column 5) and MoCo reconstruction with a motion model generated from CNN enhanced images (column 6). 3D FDK reconstruction using all projections (column 1) is provided for reference. The tumor region is indicated by red contours and a close-up is provided in transverse view. Yellow dashed lines and red arrows are used to indicate the diaphragm position to show the maintenance of temporal resolution for different methods. The display window is  $[-0.002, 0.028] \text{ mm}^{-1}$



**FIGURE 5** Reconstruction results of end-exhale phase of patient 2 from in-vivo testing set. The tumor region is indicated by red contours and a close-up is provided in coronal view. Yellow dashed lines and red arrows are used to indicate the diaphragm position. The display window is  $[-0.002, 0.028] \text{ mm}^{-1}$

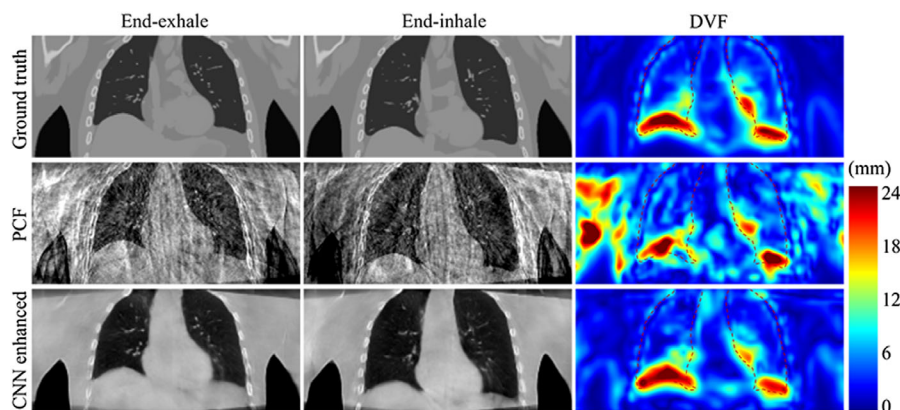
as well as ROOSTER, which indicated the advantages of this proposed algorithm over only using CNN as a post-processing procedure or directly performing MoCo reconstruction with conventional 4D-CBCT images.

### 3.2 | Phantom experiments

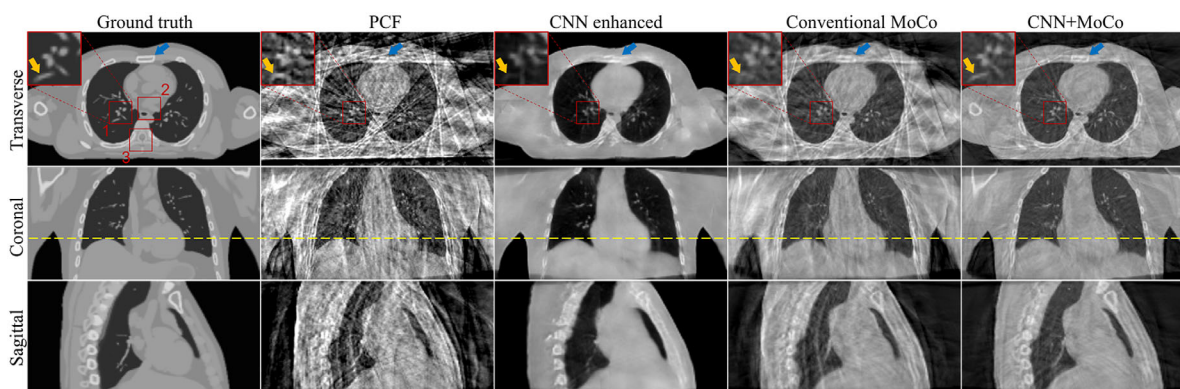
To show how the artifact-reduced images have improved the motion model estimation, Figure 6 presents the qualitative assessment of DVFs between the end-inhale phase and the end-exhale phase extracted using initial digital phantoms as ground truth, PCF images and CNN enhanced images. It should be noted that groupwise 4D registration<sup>17</sup> was utilized for the estimation of DVFs in this study such that all phase images were simultaneously registered to the same fixed image. Here, the end-exhale phase was used as fixed image and only

the coronal-view DVFs between the end-inhale phase and the end-exhale phase were shown. Compared to PCF, the DVF generated from CNN enhanced images appeared to be more similar to the ground truth. Inside the lung region (indicated by red contours), PCF generated a non-smooth DVF that was not physically realistic. Additionally, there also existed obvious noise outside the lung region for PCF, which was inconsistent with the ground truth.

Figure 7 shows the reconstruction results of XCAT phantom at end-exhale phase with ground truth. For quantitative evaluation, three regions of interest (ROIs) indicated by red squares as well as the whole region of body were chosen to compute the RMSE and NCC. The diaphragm motion was not affected in each method (see the yellow dashed line). The streaking artifacts from PCF were substantially reduced by the pre-trained CNN, however, some tiny vessel structures were missing (see



**FIGURE 6** Qualitative illustration of the magnitude of estimated DVFs ( $\|DVF(x, y, z)\|_2$ ) between end-inhale phase and the end-exhale phase using initial XCAT digital phantoms (row 1), PCF images (row 2) and CNN enhanced images (row 3). Lung regions are indicated by red contours. The display window is  $[-0.002, 0.028] \text{ mm}^{-1}$  for CBCT images and  $[0, 24] \text{ mm}$  for DVFs



**FIGURE 7** Reconstruction results of end-exhale phase of XCAT phantom. Three ROIs indicated by red squares and the whole-body region are used for quantitative evaluation. ROI 1 is zoomed in and placed on the upper left corner in transverse view. Yellow dashed line is used to indicate the diaphragm position. The display window is  $[-0.002, 0.028] \text{ mm}^{-1}$

**TABLE 1** Results of root-mean-square-error (RMSE) and normalized cross-correlation (NCC) in three ROIs and the whole body of XCAT phantom using different methods. The best results are highlighted by bold font

	ROI 1		ROI 2		ROI 3		Whole body	
	RMSE ( $\text{mm}^{-1}$ )	NCC	RMSE ( $\text{mm}^{-1}$ )	NCC	RMSE ( $\text{mm}^{-1}$ )	NCC	RMSE ( $\text{mm}^{-1}$ )	NCC
PCF	0.0041	0.73	0.0041	0.37	0.0046	0.57	0.0054	0.68
CNN enhanced	0.0050	0.89	0.0026	0.91	0.0021	0.84	0.0034	0.92
Conventional MoCo	0.0023	0.86	0.0020	0.92	0.0025	0.73	0.0032	0.85
CNN+MoCo	<b>0.0017</b>	<b>0.93</b>	<b>0.0018</b>	<b>0.94</b>	<b>0.0016</b>	<b>0.88</b>	<b>0.0020</b>	<b>0.95</b>

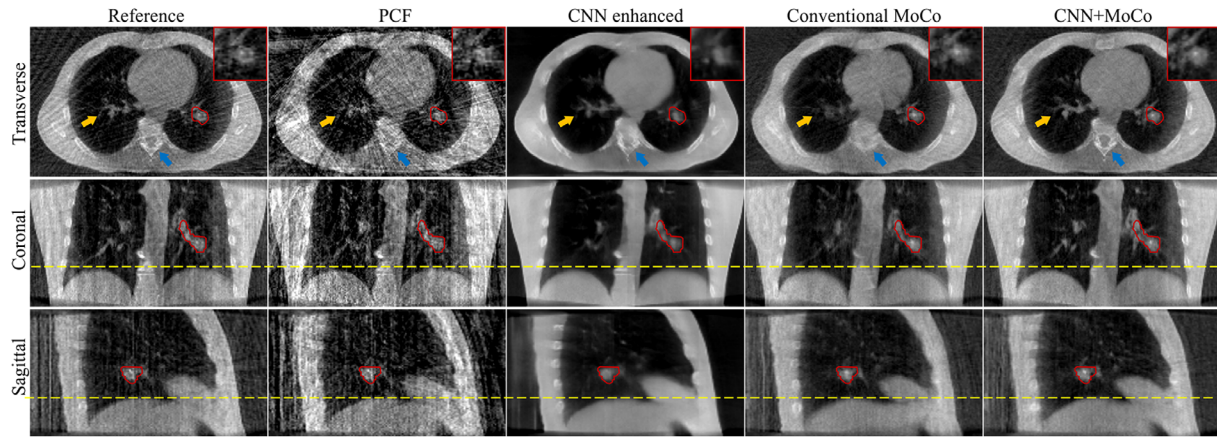
yellow arrows), and some bone structures were distorted (see blue arrows). Conventional MoCo failed to recover the lung vessels either. Although slight streaking artifacts can be observed due to the back-projection operation, these detailed structures were recovered in CNN+MoCo.

The quantitative evaluation results of different methods are shown in Table 1. CNN-based methods outperformed the conventional methods in each region. Except for ROI 2, only utilizing CNN as post-processing

can obtain better RMSE and NCC than either PCF or conventional MoCo. CNN+MoCo can further have substantial improvements for every region.

### 3.3 | SPARE challenge experiments

Similar observations held true for SPARE challenge data. Figure 8 shows the reconstructed end-exhale phase image of one typical Varian case (CV\_P4\_T\_01)



**FIGURE 8** Reconstructed end-exhale phase image of one Varian case from SPARE challenge dataset. The PTV is indicated by red contours and a close-up is provided in transverse view. Yellow dashed lines are used to indicate the diaphragm position. The display window is  $[-0.002, 0.028] \text{ mm}^{-1}$

**TABLE 2** Averaged ( $\pm$  standard deviation) RMSE and NCC in different regions over the Varian subset from the SPARE challenge dataset using different methods. The best results are highlighted by bold font

	Body		Lung		PTV	
	RMSE ( $\cdot 10^{-3} \text{ mm}^{-1}$ )	NCC	RMSE ( $\cdot 10^{-3} \text{ mm}^{-1}$ )	NCC	RMSE ( $\cdot 10^{-3} \text{ mm}^{-1}$ )	NCC
PCF	$4.0 \pm 0.9$	$0.84 \pm 0.05$	$3.8 \pm 0.8$	$0.69 \pm 0.05$	$4.0 \pm 1.0$	$0.68 \pm 0.15$
CNN enhanced	$2.9 \pm 0.2$	$0.91 \pm 0.05$	$3.2 \pm 0.6$	$0.78 \pm 0.04$	$3.1 \pm 0.6$	$0.77 \pm 0.09$
Conventional MoCo	$2.4 \pm 0.3$	$0.91 \pm 0.05$	$2.4 \pm 0.4$	$0.77 \pm 0.04$	$2.7 \pm 0.6$	$0.75 \pm 0.14$
CNN+MoCo	<b><math>2.1 \pm 0.3</math></b>	<b><math>0.93 \pm 0.04</math></b>	<b><math>2.2 \pm 0.3</math></b>	<b><math>0.81 \pm 0.06</math></b>	<b><math>2.4 \pm 0.4</math></b>	<b><math>0.79 \pm 0.10</math></b>

from SPARE challenge dataset. Since the downsampling strategy utilized for SPARE dataset did not really mimic the actual 1-minute scan, the angular spacing of projections in each respiratory bin was relatively smaller, and artifacts of PCF were not as severe as that of in-vivo dataset or the phantom dataset. CNN only mitigated the streaking artifacts from PCF, while PTV was over-smoothed (see red contours and the zoomed-in images). Conventional MoCo also suppressed the streaking artifacts, but the lung vessels (indicated by yellow arrows) and spine region (indicated by blue arrows) were distorted. Such effects were corrected in CNN+MoCo, resulting in a similar image quality as the reference image. The reconstruction results of other four Varian cases from SPARE challenge are presented in the Supplementary Material (Figures S1-S4).

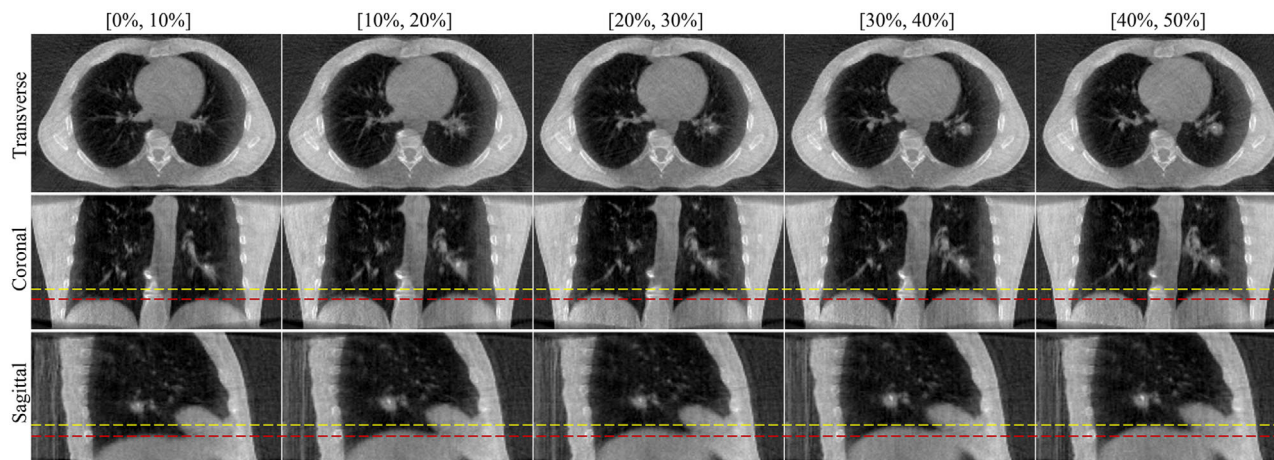
Following the evaluation procedure of SPARE challenge,<sup>23</sup> quantitative evaluations were performed for patient body, lungs and PTV respectively. The averaged results over the 5 Varian cases from the SPARE challenge dataset are shown in Table 2. Compared to PCF, both CNN enhanced and conventional MoCo had achieved better results regarding RMSE and NCC in all regions. In addition, Wilcoxon signed-rank test was used to evaluate the statistical significance of RMSE and NCC difference between the proposed CNN+MoCo and

either CNN enhanced or conventional MoCo. The RMSE of CNN+MoCo was significantly reduced compared to CNN enhanced ( $P < 0.001$ ) and conventional MoCo ( $P < 0.01$ ). The NCC of CNN+MoCo was significantly increased compared to CNN enhanced ( $P < 0.05$ ) and conventional MoCo ( $P < 0.01$ ). The separate quantitative evaluation results for each SPARE dataset are available in Tables S1-S5 in the Supplementary Material.

Figure 9 shows five CNN+MoCo reconstructions from the end-inhale to the end-exhale phase. A smooth breathing motion can be observed from this sequence of images without streaking artifacts, which demonstrates that our proposed CNN+MoCo can be performed to improve the image quality of every breathing phase while maintaining the motion information. The corresponding videos for three view planes showing the full breathing cycle and comparison with other methods are available in the Supplementary Material (Movie S1-S3).

## 4 | DISCUSSION

In this study, we proposed a deep learning-based framework to improve the quality of motion-compensated 4D-CBCT images by providing CNN-generated



**FIGURE 9** CNN+MoCo reconstructed images of the Varian SPARE challenge dataset (CV\_P4\_T\_01) from end-inhale [0%, 10%] (phase bin 1) to the end-exhale phase [40%, 50%] (phase bin 5). The red and yellow dashed lines are used to indicate the diaphragm position of end-inhale and end-exhale states, respectively. The display window is  $[-0.002, 0.028] \text{ mm}^{-1}$

artifact-mitigated 4D-CBCT images for data-driven motion model generation to the MoCo reconstruction. Since the PCF images suffered from severe streaking artifacts due to sparse sampling, the motion model directly estimated from PCF images was degraded (Figure 6) and the performance of MoCo reconstruction alone was not sufficient. An artifact-reduction CNN was proposed to enhance the initial 4D-CBCT images to generate a high-quality motion model for subsequent MoCo reconstruction and improve the final reconstructed 4D-CBCT images.

Due to the severely degraded quality of the initial FDK-reconstructed images, although the artifact-mitigated 4D-CBCT images provided by CNN achieved overall improvement in image quality, the performance was still limited by the over-smoothing and distortion effects which impacted the high contrast resolution of small structures. For example, blurring artifacts in tumor region and distortion for bone structures can be observed in Figure 8. Therefore, the CNN-generated 4D-CBCT images were used to estimate a high-quality motion model (Figure 6) such that performing MoCo reconstruction with the estimated motion model can facilitate the use of all available projections to obtain additional gains in final 4D-CBCT reconstruction and recover the resolution of small structures such as pulmonary vessels and airways.

The artifact-reduction CNN was trained only using in-vivo patient datasets but tested using in-vivo datasets, XCAT phantom, and the public SPARE challenge dataset, which demonstrated the generalization and robustness of our network. Through the evaluation of image quality using the public SPARE challenge dataset and the same quantitative metrics, our method can be compared to some previous studies. Overall, our performance on the SPARE challenge dataset was state of the art. Average RMSE of  $0.0021 \text{ mm}^{-1}$  and NCC of 0.93

were achieved in the body region using our proposed CNN+MoCo, which were  $0.0033 \text{ mm}^{-1}$  and 0.93 for the work of Huang et al.<sup>22</sup> and  $0.0022 \text{ mm}^{-1}$  and 0.79 for the work of Madesta et al.<sup>21</sup>

One limitation of our study is that the reasonable initial PCF images were required for the proposed framework. In practice, this requirement may be challenged by modern CBCT scanners with more rapid gantry rotation, such as Varian's Halcyon and Ethos platforms, since there might be too few respiratory cycles in one acquisition and the projection angular spacing for each phase bin would be too large to obtain a reasonable reconstruction. Extending our motion compensation methods for such situations will be of interest for future studies. Additionally, although this study enabled us to compensate the inter-phase breathing motion and reduce the streaking artifacts for 4D-CBCT images, the residual motion remaining in each phase bin and artifacts resulted from breathing irregularity are out of consideration. The impacts of binning strategy and breathing irregularity are the subjects of future work as ongoing collection of additional clinical datasets will be required.

We mainly focused on the improvement of MoCo reconstruction using CNN-enhanced initial 4D-CBCT images in this study. A data-driven motion modeling using groupwise registration and motion-compensated FDK, which had also shown promising performance in the SPARE challenge, were utilized for the respiratory motion compensation. The proposed framework is also applicable to other MoCo methods which use different registration algorithms for the motion modeling. Additionally, for the artifact-reduction CNN, a 3D U-Net paired with patch-wise training was utilized in our study. However, the proposed framework does not require the specific type of network. Further exploration of the performance of other network architectures will be deferred to the future.

## 5 | CONCLUSIONS

This study demonstrated the feasibility of using a CNN to provide high-quality initial 4D-CBCT images for accurate data-driven motion modeling and improve the quality of final 4D-CBCT images using a MoCo reconstruction. The pre-trained CNN effectively mitigated the streaking artifacts from conventional 4D reconstruction. By estimating an accuracy motion model using high-quality initial 4D-CBCT images, MoCo reconstruction can substantially enhance 4D-CBCT image quality through recovery of the resolution of small structures. The performance of the proposed procedure was validated by in-vivo patient datasets, XCAT phantom dataset and the publicly available SPARE challenge dataset.

## ACKNOWLEDGEMENTS

This work was partially supported by National Institute of Biomedical Imaging and Bioengineering (NIBIB) grant R03-EB028427 and National Heart, Lung, and Blood Institute (NHLBI) grant R01-HL148210. The authors also thank Dr. Matthew J. Riblett for providing help and discussions in the extraction of respiratory signals.

## CONFLICT OF INTEREST

Washington University receives research support from Varian Medical Systems, Siemens, ViewRay, and Mevion unrelated to the present work.

## REFERENCES

- Jaffray DA, Siewerdsen JH, Wong JW, Martinez AA. Flat-panel cone-beam computed tomography for image-guided radiation therapy. *Int J Radiat Oncol Biol Phys*. 2002;53(5):1337-1349. doi:10.1016/S0360-3016(02)02884-5
- Boda-Heggemann J, Lohr F, Wenz F, Flentje M, Guckenberger M. kV cone-beam CT-based IGRT. *Strahlentherapie und Onkol*. 2011;187(5):284-291. doi:10.1007/s00066-011-2236-4
- Keall PJ, Mageras GS, Balter JM, et al. The management of respiratory motion in radiation oncology report of AAPM Task Group 76. *Med Phys*. 2006;33(10):3874-3900. doi:10.1118/1.2349696
- Sonke JJ, Zijp L, Remeijer P, van Herk M. Respiratory correlated cone beam CT. *Med Phys*. 2005;32(4):1176-1186. doi:10.1118/1.1869074
- Feldkamp LA, Davis LC, Kress JW. Practical cone-beam algorithm. *J Opt Soc Am A*. 1984;1(6):612-619. doi:10.1364/JOSAA.1.000612
- Bissonnette JP, Franks KN, Purdie TG, et al. Quantifying interfraction and intrafraction tumor motion in lung stereotactic body radiotherapy using respiration-correlated cone beam computed tomography. *Int J Radiat Oncol Biol Phys*. 2009;75(3):688-695. doi:10.1016/j.ijrobp.2008.11.066
- Li T, Xing L, Munro P, et al. Four-dimensional cone-beam computed tomography using an on-board imager. *Med Phys*. 2006;33(10):3825-3833. doi:10.1118/1.2349692
- Li T, Xing L. Optimizing 4D cone-beam CT acquisition protocol for external beam radiotherapy. *Int J Radiat Oncol Biol Phys*. 2007;67(4):1211-1219. doi:10.1016/j.ijrobp.2006.10.024
- Sidky EY, Pan X. Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. *Phys Med Biol*. 2008;53(17):4777.
- Leng S, Tang J, Zambelli J, Nett B, Tolakanahalli R, Chen GH. High temporal resolution and streak-free four-dimensional cone-beam computed tomography. *Phys Med Biol*. 2008;53(20):5653-5673. doi:10.1088/0031-9155/53/20/006
- Bergner F, Berkus T, Oelhafen M, et al. An investigation of 4D cone-beam CT algorithms for slowly rotating scanners. *Med Phys*. 2010;37(9):5044-5053.
- Mory C, Auvray V, Zhang B, et al. Cardiac C-arm computed tomography using a 3D+ time ROI reconstruction method with spatial and temporal regularization. *Med Phys*. 2014; 41(2): 21903.
- Li T, Schreiber E, Yang Y, Xing L. Motion correction for improved target localization with on-board cone-beam computed tomography. *Phys Med Biol*. 2005;51(2):253-267. doi:10.1088/0031-9155/51/2/005
- Rit S, Wolthaus JWH, van Herk M, Sonke JJ. On-the-fly motion-compensated cone-beam CT using an a priori model of the respiratory motion. *Med Phys*. 2009;36(6):2283-2296. doi:10.1118/1.3115691
- Brehm M, Paysan P, Oelhafen M, Kachelrieß M. Artifact-resistant motion estimation with a patient-specific artifact model for motion-compensated cone-beam CT. *Med Phys*. 2013;40(10):1-13. doi:10.1118/1.4820537
- Wang J, Gu X. Simultaneous motion estimation and image reconstruction (SMEIR) for 4D cone-beam CT. *Med Phys*. 2013;40(10):101912.
- Riblett MJ, Christensen GE, Weiss E, Hugo GD. Data-driven respiratory motion compensation for four-dimensional cone-beam computed tomography (4D-CBCT) using groupwise deformable registration. *Med Phys*. 2018;45(10):4471-4482. doi:10.1002/mp.13133
- Ahishakiye E, Bastiaan Van Gijzen M, Tumwiine J, Wario R, Obungoloch J. A survey on deep learning in medical image reconstruction. *Intell Med*. 2021;1(03):118-127.
- Zhi S, Duan J, Cai J, Mou X. Artifacts reduction method for phase-resolved Cone-Beam CT (CBCT) images via a prior-guided CNN. *Medical Imaging 2019: Physics of Medical Imaging*. SPIE; 2019:562-567.
- Jiang Z, Chen Y, Zhang Y, Ge Y, Yin FF, Ren L. Augmentation of CBCT reconstructed from under-sampled projections using deep learning. *IEEE Trans Med Imaging*. 2019;38(11):2705-2715.
- Madesta F, Sentker T, Gauer T, Werner R. Self-contained deep learning-based boosting of 4D cone-beam CT reconstruction. *Med Phys*. 2020;47(11):5619-5631. doi:10.1002/mp.14441
- Huang X, Zhang Y, Chen L, Wang J. U-net-based deformation vector field estimation for motion-compensated 4D-CBCT reconstruction. *Med Phys*. 2020;47(7):3000-3012.
- Shieh CC, Gonzalez Y, Li B, et al. SPARE: sparse-view reconstruction challenge for 4D cone-beam CT from a 1-min scan. *Med Phys*. 2019;46(9):3799-3811. doi:10.1002/mp.13687
- Rit S, Oliva MV, Brousmiche S, Labarbe R, Sarrut D, Sharp GC. The Reconstruction Toolkit (RTK), an open-source cone-beam CT reconstruction toolkit based on the Insight Toolkit (ITK). *J Phys Conf Ser*. 2014;489:12079. doi:10.1088/1742-6596/489/1/012079
- Chen H, Zhang Y, Kalra MK, et al. Low-Dose CT with a residual encoder-decoder convolutional neural network. *IEEE Trans Med Imaging*. 2017;36(12):2524-2535. doi:10.1109/TMI.2017.2715284
- Schulze R, Heil U, Gross D, et al. Artefacts in CBCT: a review. *Dentomaxillofac Radiol*. 2011;40(5):265-273. doi:10.1259/dmfr/30642039
- Çiçek Ö, Abdulkadir A, Lienkamp S, Brox T, Ronneberger O. 3D U-Net: learning dense volumetric segmentation from sparse annotation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2016:424-432.
- Odena A, Dumoulin V, Olah C. Deconvolution and checkerboard artifacts. *Distill*. 2016;1(10):e3.

29. Zhang K, Zuo W, Chen Y, Meng D, Zhang L. Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising. *IEEE Trans Image Process*. 2017;26(7):3142-3155.
30. Klein S, Staring M, Murphy K, Viergever MA, Pluim JPW. elastix: a toolbox for intensity-based medical image registration. *IEEE Trans Med Imaging*. 2010;29(1):196-205. doi:10.1109/TMI.2009.2035616
31. Schafer D, Borgert J, Rasche V, Grass M. Motion-compensated and gated cone beam filtered back-projection for 3-D rotational X-ray angiography. *IEEE Trans Med Imaging*. 2006;25(7):898-906.
32. Zijp L, Sonke JJ, Herk M. Extraction of the respiratory signal from sequential thorax cone-beam x-ray images. *International Conference on the Use of Computers in Radiation Therapy*. 2004;507-509.
33. Cooper BJ, O'Brien RT, Balik S, Hugo GD, Keall PJ. Respiratory triggered 4D cone-beam computed tomography: a novel method to reduce imaging dose. *Med Phys*. 2013;40(4):41901.
34. Segars WP, Sturgeon G, Mendonca S, Grimes J, Tsui BMW. 4D XCAT phantom for multimodality imaging research. *Med Phys*. 2010;37(9):4902-4915. doi:10.1118/1.3480985
35. Joseph PM. An improved algorithm for reprojecting rays through pixel images. *IEEE Trans Med Imaging*. 1982;1(3):192-196. doi:10.1109/TMI.1982.4307572
36. La Riviere PJ, Billmire DM. Reduction of noise-induced streak artifacts in X-ray computed tomography through spline-based penalized-likelihood sinogram smoothing. *IEEE Trans Med Imaging*. 2005;24(1):105-111.
37. Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. JMLR Workshop and Conference Proceedings; 2010:249-256.
38. Kingma DP, Ba J. Adam: A method for stochastic optimization. *arXiv Prepr arXiv14126980*. Published online 2014.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Zhang Z, Liu J, Yang D, Kamilov US, Hugo GD. Deep learning-based motion compensation for four-dimensional cone-beam computed tomography (4D-CBCT) reconstruction. *Med Phys*. 2023;50:808–820. <https://doi.org/10.1002/mp.16103>