IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 13, 2019, accepted February 7, 2019, date of publication February 14, 2019, date of current version April 3, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2899385

Convolutional Neural Network (CNN) Based Three Dimensional Tumor Localization Using Single X-Ray Projection

RAN WEI[1], FUGEN ZHOU[1,2], BO LIU[1,2], XIANGZHI BAI[1,2], DONGSHAN FU[2], YONGBAO LI[3], BIN LIANG[4], AND QIUWEN WU[5]

[1]Image Processing Center, Beihang University, Beijing 100191, China
[2]Beijing Advanced Innovation Center for Biomedical Engineering, Beihang University, Beijing 100083, China
[3]State Key Laboratory of Oncology in South China, Sun Yat-Sen University Cancer Center, Collaborative Innovation Center for Cancer Medicine, Department of Radiation Oncology, Guangzhou 510060, China
[4]Department of Radiation Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China
[5]Department of Radiation Oncology, Duke University Medical Center, Durham, NC 27710, USA

Corresponding authors: Bo Liu (bo.liu@buaa.edu.cn) and Xiangzhi Bai (jackybxz@buaa.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61601012, Grant 81801799, and Grant U1736217, and in part by the National Key R&D Program of China under Grant 2017YFC0113100.

ABSTRACT Accurate localization of lung tumor in real time based on a single X-ray projection is of great interest to the tumor-tracking radiotherapy but is very challenging. In this paper, a convolutional neural network (CNN)-based tumor localization method was proposed to address this problem with the aid of principal component analysis-based motion modeling. A CNN regression model was trained before treatment to recover the ill-conditioned nonlinear mapping from the single X-ray projection to the tumor motion. Novel intensity correction and data augmentation techniques were adopted to improve the model's robustness to the scatter and noise in the X-ray projection image. During treatment, the volumetric image and tumor position could be obtained by applying the CNN model on the acquired X-ray projection. This method was validated and compared with the other state-of-the-art methods on three real patient data. It was found that the proposed method could achieve real-time tumor localization with much higher accuracy (<1 mm) and robustness.

INDEX TERMS Convolutional neural network (CNN), PCA breathing motion modeling, single x-ray projection, tumor localization, volumetric imaging.

I. INTRODUCTION

Lung cancer has been the worldwide leading cause of cancer death [1]. In radiation therapy for lung cancer, tumor motion caused by respiration has a large impact on the efficacy of treatment [2], [3]. To improve the accuracy of beam delivery, various motion-management techniques have been proposed [4], [5]. Especially, repositioning of radiation beam to follow the tumor movement, or tumor-tracking radiotherapy, has been increasingly utilized in clinical treatment. In these techniques, the accurate tumor localization plays a vital role.

To accurately localize lung tumors, numerous approaches have been proposed over the past years [6]–[8]. These methods can be categorized basically into marker-based, surrogate-based, and methods of direct tumor localization [9]. The marker-based localization is currently most commonly used and proven techniques. However, the requirement of implanting fiducial markers into the body remains the major concern of this method, as the marker implanting may increase the risk of pneumothorax [10]. In addition, the accuracy may decline with time due to the marker migration [10]. Surrogate-based technique is an indirect way for tumor localization based on external signals whose variations are correlated with the tumor motion, such as skin marker or lung volume [10], [11]. However, the relationship between the surrogates and tumor may vary with time, leading to inaccurate tumor localization. In recent years, the direct tumor localization methods based on x-ray projection have gained momentum and drawn more interest. It utilizes image processing techniques to directly localize the tumor position from the image. As it is noninvasive and could be more

The associate editor coordinating the review of this manuscript and approving it for publication was Syed Muhammad Anwar.

reliable than surrogate-based methods, direct tumor localization methods have great potential for clinical use.

However, because only one x-ray projection can be acquired at each gantry angle at one time on common linear accelerator (LINAC) treatment machines [11], it is not straight forward to directly localize the tumor position without implanting markers, especially in rotational radiotherapy. Prior knowledge is usually needed for this type of methods [9], [10]. Principal component analysis (PCA) based motion modeling is one typical method based on prior knowledge [12]–[14], which could reduce the degree of freedom of the problem tremendously. Based on this model, several methods have been proposed [9]–[11], which could perform volumetric imaging from one x-ray projection and obtain the tumor position. The resulting volumetric image can be used in tumor localization, dose assessment and other clinical tasks [9]. Two types of methods have been proposed to obtain the volumetric image from single x-ray projection: registration based methods and regression based methods. The first registration based method, namely the Motion-Model (MM) method, was proposed by Li *et al.* [10], in which the 3D trajectory of tumor motion was estimated by a forward matching process between a reference 3D image and the acquired 2D x-ray projection. Then the Motion Model – Free Deformation (MM-FD) method was put forward by Zhang *et al.* [11], which utilized a free-deformation registration process to improve the accuracy of tumor localization. These methods usually tried to minimize the intensity differences between the computed digitally reconstructed radiographs (DRR) and the acquired x-ray projections in the matching process. The robustness is poor due to the intensity mismatch caused by the scatter and noise in the projection. While the linear intensity correction could be applied to alleviate this problem, it is still a major concern for most of the time and can degenerate the registration accuracy. Moreover, these methods involve the complex process of iterative optimization, which could not achieve the tumor tracking in real-time [9]. To address these issues, Xu *et al.* [9] put forward a regression based method. A linear model was built to approximate the complex nonlinear mapping from several selected informative image patches to the organ motion. Although real-time tumor tracking has been achieved in digital and physical phantom experiments, the linear model may not be adequate to describe the complex relationship between the intensity variation and breathing motion in realistic cases. Besides, the method applied piecewise linear correction for the scatter and noise that exist in acquired x-ray projection but not in computed DRR. This may not be adequate to describe real patient data.

Recently, convolutional neural network (CNN) has been applied to many areas of medical image processing. CNN could automatically extract hierarchy features from image data and has strong non-linear modeling capability, which may be potentially suitable for modeling the complex relationship between the breathing motion and single x-ray projection

In view of this, we proposed a CNN based volumetric imaging method to achieve the tumor localization in real time. In our method, a CNN regression model was utilized to approximate the complex nonlinear mapping from single x-ray projection to the tumor motion. Unlike previous methods, we utilized a novel intensity correction process for the 3D-CT images to narrow down the differences between DRR and cone-beam CT (CBCT) projections. And a data augmentation technique was applied to enhance the robustness of the CNN model to noise. As a result, the scatter and quantum noise were taken into account in the learning process instead of being handled by the intensity correction in the application process [9]–[11]. For the computation time, since the forward calculation of CNN could be highly parallelized, our method could achieve real-time tumor tracking. To evaluate the efficacy of our method, we compared it with the registration based methods using the 4D digital extended cardiac-torso (XCAT) phantom [15] and three real patient cases.

II. METHOD

As shown in Fig.1, our method consists of two stages: pre-treatment training stage and online application stage. In the training stage, a PCA based lung motion model was established from the planning 4D-CT, according to which we generated 1000 3D-CT images DRRs by randomly sampling the PCA coefficients. Next, to handle the problem of the

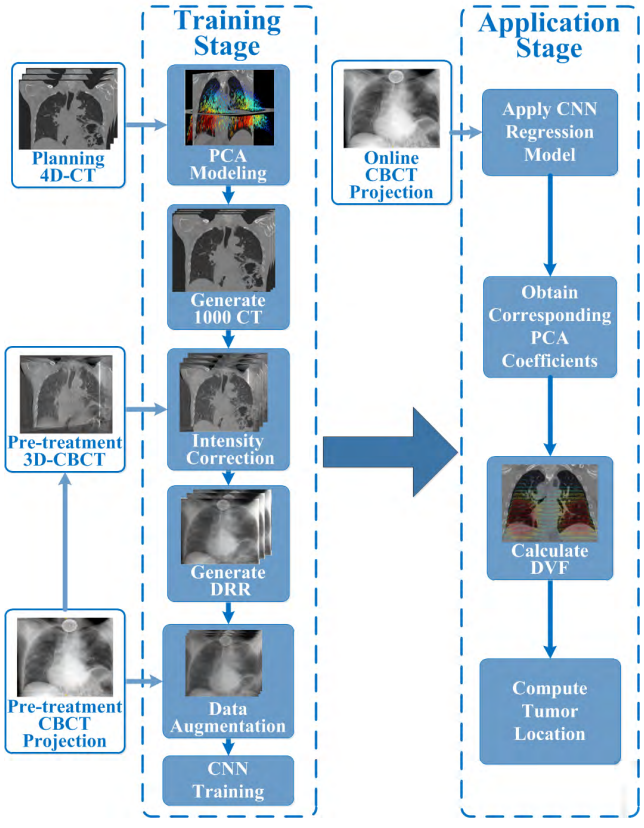

FIGURE 1. Workfow of the proposed method.

intensity discrepancy caused by scatter and noise, an intensity correction process and a data augmentation step were performed. In the intensity correction process, the 3D-CBCT acquired during pre-treatment simulation procedure is used as a reference image, the intensity differences between 3D-CT images and 3D-CBCT image is examined and 3D-CT image intensity is modified accordingly to increase the similarity between DRRs from the corrected 3D-CT and the CBCT projections. These DRRs are further augmented by adding noise whose distributions were calculated from the actual CBCT projections to enhance the robustness of the regression model. Finally, a CNN regression model with a patient-specific loss function was trained to estimate the complex nonlinear mapping from single x-ray projection to the organ motion.

In the application stage, the trained CNN regression model was applied on the online acquired single x-ray projection to determine the corresponding PCA coefficients and the deformation vector field (DVF). With the generated DVF, the volumetric image and tumor location were finally obtained.

A. GENERATION OF TRAINING DATA

The PCA-based motion model is the basis of generating training data. We describe the model in the following.

Consider there are N phases of 3D-CT images within one breathing cycle. Without loss of generality, the volumetric image at the end-expiration phase was selected as the reference CT image. By utilizing deformable image registration [16]-[20], a set of DVFs between this reference CT image and 3D-CT images at other phases can be obtained. Each calculated DVF corresponding to one phase was represented by a column vector of the size $[3 \times N_{voxelCT}, 1]$, where 3 represents the 3-dimensional motion; $N_{voxelCT}$ denotes the number of voxels in CT image. And a matrix M was formed by these N column vectors. Principal component analysis (PCA) was performed on the matrix M and the principal components (eigenvectors) of the lung motion were acquired. Then any DVF relative to the reference CT image could be represented by a linear combination of a group of eigenvectors as:

$$F \approx \tilde{F} + \sum_{i=1}^{k} p_i q_i. \tag{1}$$

Here, F is a DVF, and $\tilde{F}$ is the averaged motion vector field. p and q refer to the PCA eigenvectors and their corresponding PCA coefficients, respectively. The subscript i is the index of the eigenvectors. k is the number of principal components that we utilized to describe the tumor motion. Based on previous studies [16], we chose three principal components which were able to reliably represent the original DVF. Using more PCA coefficients could increase the computational complexity and may cause overfitting problem [9].

To generate the training data, we first calculated 1000 DVFs by randomly sampling the PCA coefficients. Because the range of tumor motion during treatment could be larger than that in the 4D-CT from CT simulation, the sampling range was set to be larger than the amplitudes of the PCA coefficients in the planning 4D-CT. In this study, the range of possible tumor motion was set to 5 cm, which was much larger than the averaged range of tumor motion in patients [9], [10]. Then, the corresponding volumetric images were obtained by deforming the reference CT with the generated DVFs.

The DRRs computed from the generated 3D-CT Image can be in principle used to train the CNN model directly. However, these DRRs were free of scatter or quantum noise that exists in the actual x-ray projection images. As demonstrated in the following experiments, the CNN model trained using these ideal DRRs performed poorly on realistic x-ray projections. Therefore we performed the following additional intensity correction and data augmentation.

B. INTENSITY CORRECTION FOR 3D-CT IMAGE

To realize the accurate tumor localization for clinical treatment, the intensity discrepancy between the training DRRs and the CBCT projections need to be minimized. In this work, we proposed an intensity correction method for 3D-CT images so that the computed DRRs have a high similarity with the CBCT projections. The idea was based on the fact that 3D-CBCT was reconstructed using the actual projections and the DRR of the 3D-CBCT image has great similarity with the CBCT projection. Therefore, it is possible to improve the similarity between the DRRs and the CBCT projections by correcting or replacing the intensity of the 3D-CT according to the 3D-CBCT.

Applying deformable registration between the 3D-CBCT image and the 3D-CT image is one way to perform intensity correction. However, the moving part of the patient's anatomy in the 3D-CBCT image, such as tumor and diaphragm, was blurred due to the respiratory motion. The blurry may be propagated to the corrected 3D-CT and the generated DRR. Because of this, we proposed to separate the ribcage part (with less motion) and moving tissue part from both the 3D-CT image and 3D-CBCT image, and establish the mapping for the ribcage part and moving tissue part separately. The work flow of the intensity correction is shown in Fig.2.

For the ribcage part, including the rib, spine and other parts of patient anatomy that have little movement, we applied a deformable registration between the 3D-CBCT image and 3D-CT image. Then we deformed the 3D-CBCT ribcage image with the calculated DVF, and the ribcage part of the 3D-CT image was replaced by the deformed 3D-CBCT image.

The moving tissue part extracted consisted of lung, heart, tumor, diaphragm and other tissues which had large movement caused by breathing motion. For this part, we utilized a linear intensity correction strategy:

$$I_{correctedCT} = (I_{CT} - Mean_{CT}) \cdot \frac{Std_{CBCT}}{Std_{CT}} + Mean_{CBCT}. \tag{2}$$

Here, I_{CT} and $I_{correctedCT}$ represented the intensity of the original moving tissue in 3D-CT image and the corrected moving tissue 3D-CT image, respectively. $Mean_{CT}$ and Std_{CT}

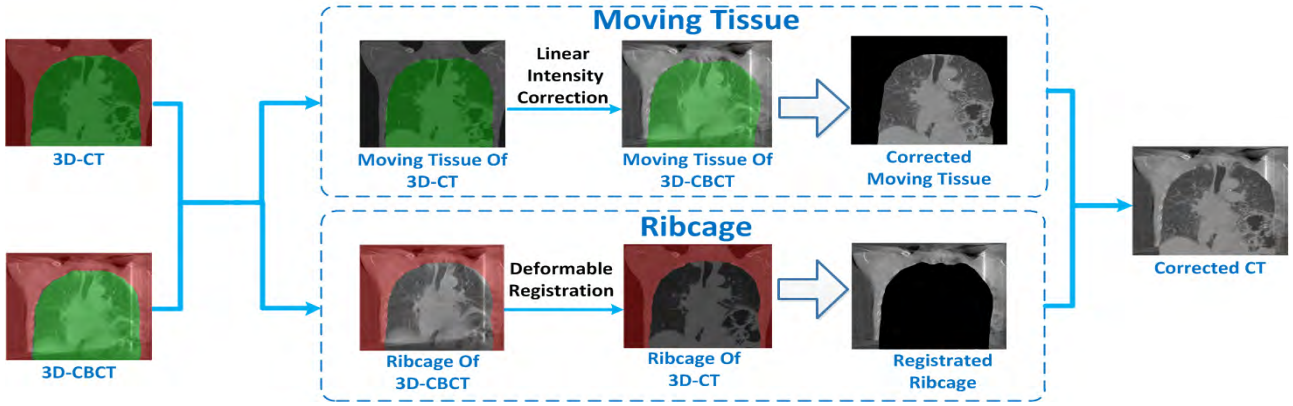

FIGURE 2. Workflow of intensity correction.

denoted the mean value and the standard deviation of intensity of the original moving tissue in 3D-CT image, $Mean_{CBCT}$ and Std_{CBCT} for the corresponding 3D-CBCT image. This way, the corrected 3D-CT moving tissue image should have the same mean value and standard deviation of intensity with the 3D-CBCT.

Finally, the whole corrected 3D-CT image is obtained by combining the ribcage part and corrected moving tissue. As Shown in Fig. 3, compared to the DRR from uncorrected CT, the DRR from corrected CT has much greater similarity with the CBCT projection.

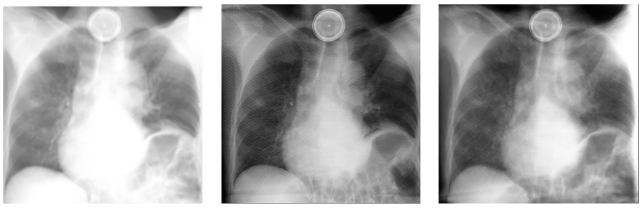

FIGURE 3. Comparison among (a): DRR of uncorrected CT. (b): CBCT projection. (c): DRR of corrected CT.

C. DATA AUGMENTATION

According to Fig. 3, although the differences between the DRR of corrected CT and CBCT projection had been significantly reduced, there still exists intensity inconsistency between the DRR and CBCT projection, which would hamper the accuracy in tumor localization. To circumvent this, we performed some experiments in the following manner. We calculated the residual images between the CBCT projections with different breathing phases and the corresponding DRRs of the corrected CT. It was found that the joint two-dimensional histograms between the DRR and residual images had basically the same probability distribution for the intensity, which meant the differences between the DRR and CBCT projections could be treated as a sort of noise with certain distribution. Based on these findings, we proposed a data augmentation process by adding random noise to DRRs with the distribution calculated from the two dimensional histogram to enhance the robustness of the CNN regression model.

The noise was chosen to follow a conditional probability distribution and determined as follows. Firstly, we calculated the residual images between the CBCT projection and its corresponding DRR generated from corrected 3D-CT. The intensity of residual image and DRR from corrected CT image was rescaled into integer range [0, 1000].

Then, the joint histogram between the DRR image and residual image was calculated as follows:

$$\mathrm{Hist}(i,j) = \frac{\sum_{x=1}^{W}\sum_{y=1}^{H} \delta[I_{DRR}(x,y)-i]\cdot\delta[I_{Res}(x,y)-j]}{W\cdot H}. \quad (3)$$

Here, Hist(i, j) denotes the probability of the intensify of DRR being i and the intensity of residual image being j simultaneously. W is the width of the image in pixels, and H is the height of the image. $I_{DRR}(x,y)$ and $I_{Res}(x,y)$ denote the intensity of DRR and residual image at the coordinate (x,y), respectively. δ is the discrete impulse function.

According to the above histogram, we generated random noise image whose intensity followed a conditional probability distribution as below:

$$\mathrm{P}(I_{Noise}(x,y)=n|I_{DRR}(x,y)=m) = \frac{\mathrm{Hist}(I_{DRR}(x,y),n)}{\sum_{k=1}^{1000}\mathrm{Hist}(I_{DRR}(x,y),k)}. \quad (4)$$

Here, $I_{DRR}(x,y)$ denotes the intensity of the pixel in rescaled DRR at the coordinate (x,y). $I_{Noise}(x,y)$ indicates the value of the random noise image at the same coordinate. Hist() was the two dimensional histogram we just obtained. P($Noise(x,y) = n|I_{DRR}(x,y)$) means the probability of $I_{Noise}(x,y)$ being the value of n on the condition that I_{DRR}(x, y) equals to m. Then, m was rescaled to the original range in residual image.

Finally, we could obtain the final training projection by integrating the calculated DRR with the above random noise

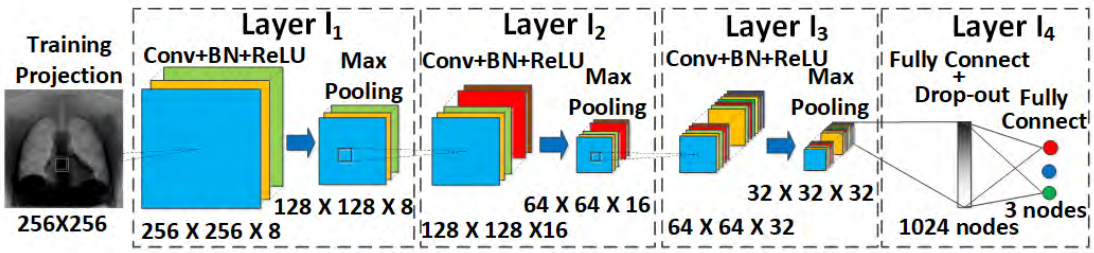

FIGURE 4. Architecture of CNN regression model.

image, as follows:

$$I_{FinalProj} = I_{DRR} + W_{random} \cdot I_{Noise}. \tag{5}$$

Here, $I_{FinalProj}$ is the final training projections. I_{DRR} stands for the intensity of the converted DRR. W_{random} is a random coefficients matrix, whose each element is a random number with the continuous uniform distribution in the interval of [0, 1). I_{Noise} was the random noise we generated following Eq. (4). For each DRR, we generated 100 training projections to enhance the robustness of the trained CNN model.

D. CNN REGRESSION MODEL

The nonlinear mapping from single x-ray projection to the tumor motion is complex and strong non-linear modeling capability is required for the network. Meanwhile, the structure of the network should not be too complicated, since it needs to be forward calculated in real-time. Based on these considerations, the structure of the CNN and training process were designed as in Fig. 4.

1) NETWORK ARCHITECTURE

The proposed CNN consists of four layers, as shown in Fig. 4. The first three layers are aimed at extracting hierarchical features from the input image and each one contains two sub-layers:

(1) Conv + BN + ReLU: for this type of sub-layer in Layer l_i ($i = 1, 2, 3$), a 3×3 convolution calculation was firstly conducted to generate 2(i+2) feature maps. Then the rectified linear units (ReLU) were applied for nonlinearity. Batch normalization was added between the convolution and the ReLU to speed-up the training.
(2) Max-pooling: A 2×2 max-pooling with stride 2 was utilized to down-sample the feature map generated from the previous sub-layer.

Layer l_4 was a fully connected multi-layer network and composed of three layers: a fully connected layer of 1024 units, a drop-out layer with ratio of 0.5 and an output layer with 3 neurons, corresponding to the estimated PCA coefficients.

In this CNN regression model, three convolutional layers were selected based on experimental results. When only two convolutional layers were included in the network, it seemed that the network was insufficient to model the complex nonlinear mapping from single x-ray projection to tumor motion. The training loss could not converge to the stopping threshold, even if we increased the number of the feature maps extracted by the convolutional layers. On the other hand, we found more than three convolution layers in the network would lead to over-fitting problem. The experiment showed that four convolutional layers led to a faster convergence of the training loss but a lower accuracy in estimating tumor location on the testing dataset. Therefore, we utilized three convolution layers in this network, so that the model was able to recover the complex nonlinear mapping accurately.

2) LOSS FUNCTION

To have a fixed criterion of stopping training for different patients, we utilized a normalized mean square error as the loss function. Moreover, to improve the accuracy of tumor localization, weights of PCA coefficients were introduced to the loss function, which ensured the first coefficient had the highest estimation accuracy. The expression of the function given as follows:

$$H = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{3} \cdot \left\| u \circ w_{coeff} \circ (y_i - G(x_i, W)) \right\|_2. \tag{6}$$

Here N stands for the number of training samples in one iteration, which was 200 in our experiment. $||. ||_2$ denotes the L_2 norm, and $\circ$ means the element-wise multiplication of matrices. x_i is the i-th training image, and y_i refers to the PCA coefficients, namely the label of the i-th training image. W is the parameters of the network which need to be learned. $G(x_i, W)$ is the output of the regression model. $u = [1/u_1, 1/u_2, 1/u_3]$ is the normalization factor, where u_i is the standard deviation of the i-th PCA coefficients in training labels. By normalizing the PCA coefficients, we can have the same criterion of stopping the training for different patients. In our experiment, the training was stopped when the value of the loss function was less than 0.05, which meant the average error of estimated PCA coefficients was less than 5%. w_{coeff} represents the weights of the PCA coefficients, which is a unit vector. In our method, w_{coeff} was empirically set to be $[2/\sqrt{6}, 1/\sqrt{6}, 1/\sqrt{6}]$ based on experiments. The loss function was optimized using Adam method with exponentially decaying learning rate.

3) TRAINING

During the training, the PCA coefficients were used as the labels, and all the 100000 training images were divided into 1000 groups. The images in one group shared the same PCA coefficient, but consisted of different noise information. The model was trained with the mini-batch stochastic gradient decent method. The size of the training batch was set to be 200. At each training iteration, 200 groups were first randomly selected from the 1000 groups. Then, one training image was randomly picked from each chosen group to form the final training batch, which was fed into the model along with their corresponding labels.

E. ONLINE TUMOR LOCALIZATION

To localize the tumor position, we need to find out the tumor localization in reference CT image first. Once an x-ray projection is acquired during the treatment, it will be fed into the CNN regression model to obtain the PCA coefficients corresponding to the x-ray projection. Then the DVF of the patient anatomy could be obtained with the generated PCA coefficients, as the following:

$$F_{patient} = \tilde{F} + \sum_{i=1}^{3} p_i Q_i. \tag{7}$$

Here, $F_{patient}$ is the desired DVF of the patient's anatomy. $\tilde{F}$ and p_i refer to the averaged motion vector field and the i-th principal components in the PCA motion model, respectively. Q_i is the i-th PCA coefficient we obtained from the CNN model. Since the tumor position is known in reference CT [21], [22], we could find out the tumor movement in 3D space according to the three dimensional vector of the same position in the generated DVF. The volumetric image corresponding to that x-ray projection image can be generated by deforming the reference CT with the generated DVF. This volumetric image may be used for dose assessment and other clinical tasks.

F. EXPERIMENT DESIGN

To demonstrate the efficacy of the proposed method, we tested the performance of the CNN based method on both the simulated data and actual patient images.

1) EXPERIMENTS ON SIMULATION IMAGES

a: EXPERIMENTAL DATA

To evaluate the algorithm, we generated a ten phase 4D-CT with a 10 mm-diameter tumor in the middle of left lung using the 4D XCAT phantom [15]. The CT image of each phase had a resolution of $256 \times 256 \times 151$ voxels with a physical size of $51.2 \times 51.2 \times 30.2$ cm3. Respiratory signal from actual patient was utilized, and the respiration-induced tumor motion had peak-to-peak amplitudes of 0.2 cm, 1 cm and 1.5 cm in the left-right (LR), anterior-posterior (AP) and superior-inferior (SI) directions, respectively. The 4D-CT data was used to train the network as presented previously.

To simulate the 3D-CBCT for intensity correction, the projection images including breathing motion are needed. To achieve this, four new sets of DVFs were inserted between two neighboring phases of 4DCT with cubic spline interpolation, resulting in a total of 50 sets of DVFs. They are used to generate 50 CTs representing patient anatomy during a breathing cycle of 5 seconds. A typical LINAC gantry rotates at the speed of 1 round/per minute, therefore 600 projection images would be produced during one CBCT acquisition which contains 12 breathing cycles. Each projection image is generated from one of the 50 CTs according to the phase by the EGS_CBCT program [23], which considers the scatter, quantum noise and other features that exist in realistic x-ray projections. Then the 3D-CBCT image was reconstructed using these simulated scatter-included projection images.

For the test dataset, we first obtained volumetric images by deforming the reference CT using DVFs generated through randomly sampling the PCA coefficients. By using a wide sampling range, the generated images contain tumor motion range of 0.5 cm (LR), 1.5 cm (AP) and 2 cm (SI), larger than that in 4D-CT. Then, the test x-ray projections were computed under the same geometry with the training ones using the EGS_CBCT program whose resolution were 256×256 pixels with a spacing of 2 mm. The test volumetric images and PCA coefficients were utilized as ground truth in the evaluation of the algorithm.

b: EVALUATION METRICS

We evaluate the proposed method from three aspects: the accuracy of the generated PCA coefficients, the reconstructed volumetric images and the estimated tumor location. The computation time of our algorithm will also be discussed.

To quantitatively analyze the accuracy of the estimation of PCA coefficients, a normalized error was computed as follows:

$$e_i = \frac{\|w_i - W_i\|_2}{s_i}. \tag{8}$$

Here, the subscript i ($i = 1, 2, 3$) is the index of PCA coefficients. w_i, W_i, and s_i represent the i-th estimated PCA coefficient, the i-th ground truth PCA coefficient and the standard deviation of the i-th ground truth PCA coefficient, respectively.

We also compared the ground truth volumetric images and the predicted volumetric images using the normalized *RMSE*, as given below:

$$RMSE_{norm} = \frac{1}{u}\sqrt{\frac{\sum_{i=1}^{n}(x(i) - X(i))^2}{n}}. \tag{9}$$

Here, x refers to the intensity of the calculated images. X means the intensity of the ground truth images. n represents the number of voxels in the image. u is the standard deviation of the intensity of ground truth images.

To evaluate the accuracy of tumor localization of our method, we computed the average relative errors between the

TABLE 1. Tumor parameters of real patients.

Case	Size (cm^3)	Motion range in SI (mm)
Patient 1	1.4	18
Patient 2	1.5	13
Patient 3	25.5	10

estimated tumor position and ground truth in three directions, as given below:

$$\overrightarrow{P}_{error} = \frac{\sum \left| \overrightarrow{P}_{estimated} - \overrightarrow{P}_{ground} \right|}{N}. \quad (10)$$

Here, $\overrightarrow{P}_{error}$ is a three dimensional vector that represent the average relative errors. $\overrightarrow{P}_{\text{estimated}}$ infers to the vector of estimated tumor position in LR, AP and SI directions. $\overrightarrow{P}_{ground}$ is the ground truth of the tumor location. N refers to the number of the test subjects, which was 100 in our experiment.

To show the advantage of the proposed method, we also compared it with two conventional registration-based methods: the MM method of Li *et al.* [10] and MM-FD method of Zhang *et al.* [11]. Linear intensity correction step was utilized to handle the inconsistency between the DRR and acquired x-ray projections for the MM method and MM-FD method.

2) EXPERIMENTS ON REAL PATIENT STUDY

a: PATIENT DATA

The algorithm was also evaluated on three patient datasets. For the first patient, the 4D-CT was acquired on four-slice GE LightSpeed CT scanner and synchronized with the Varian RPM system. The CT dimensions was 512 × 512 × 106 voxels with a resolution of 1.26 × 1.26 × 3 mm^3. The cone beam projections were acquired for treatment positioning on an Elekta VersaHD LINAC using half-fan mode (120 kV, 20 mA, and 20 ms). The projection had a dimension of 512 × 512 pixels and a resolution of 0.8 × 0.8 mm^2. Both the second and third patient data were from RTK website (http://wiki.openrtk.org/index.php/RTK/4DROOSTERReconstruction). The 4D-CT datasets were 512 × 512 × 170 voxels with a resolution of 0.88 × 0.88 × 2 mm^3 and 512 × 512 × 218 voxels with a resolution of 1.37 × 1.37 × 2 mm^3, respectively. The CBCT projections of these two patients were both 512 × 512 pixels with a resolution of 0.8 × 0.8 mm^2. For the three patients, the size and motion range of the tumors were shown in Table 1.

For each patient, the CBCT projections at angle of 0°, 30°, 45°, 60° and 90° were utilized to test the performance of the proposed method. And the CBCT projections at other angles were utilized to reconstruct the 3D-CBCT image. With the 4D-CT data and reconstructed 3D-CBCT, the CNN regression models were trained as described previously.

b: EVALUATION METRICS

For our experiment, we examined retrospectively some patient data. No patient was implanted with fiducial markers, therefore real-time location of tumor could not be automatically obtained to evaluate our algorithm. However, the tumors are visible on some of the cone beam projections. Therefore, for each projection where tumor was visible, the contour of the tumor was defined by one clinician and confirmed by another. For the first two patients, the size of the tumor were small and isocenter of the cone beam scan was not placed at the center of tumor, therefore the tumors were only visible at angles ranging from 30° to 95°. For the third patient, the tumor was fairly large and it was basically visible at angle from 10° to 120°. Based on the angle ranges where the tumors were visible for all three patients, we chose projections at 30°, 45°, 60°, and 90° to test the efficacy of our method. To reduce the uncertainty of the true position of tumor, its contours on the chosen projections were further refined based on the tumor contours defined in adjacent projections, this ensured the continuity of the tumor motion. The tumor centroid positions were calculated from the refined clinician-defined tumor contours on the chosen projections and considered as the ground truth. Then, we projected the estimated tumors' 3D location onto the 2D imager, and the errors between estimated tumor position in 2D image and the ground truth were computed in horizontal and vertical directions to quantitatively evaluate the performance of the proposed method. Moreover, the result of MM and MM-FD methods were also included in the comparison.

We also tested our method on the projections where tumors were barely visible. For the CBCT projections at 0° where tumor cannot be observed, the diaphragm position was utilized for the evaluation of the proposed method. Similarly, the diaphragm position was marked by one clinician and confirmed by another. Then, we calculated the DRR of the reconstructed volumetric image, and performed similar evaluation and comparisons with the MM and MM-FD methods.

It is worth to point out that the uncertainty in the ground truth of tumor position cannot be avoided, even though some measures had been taken to minimize it. Therefore, we divided four clinician into two groups and estimated the uncertainty in the ground truth definition based on the difference between these two groups. It was shown that the errors in centroid position calculated from clinician-defined tumor contours were about 0.5 – 1.0 mm on average. And the error in the diaphragm height was around 0.6 - 0.8 mm

III. RESULTS

A. RESULTS FOR SYNTHETIC STUDY

1) PCA COEFFICIENTS

In this section, we compared the calculated PCA coefficients and the ground truth PCA coefficients. The normalized relative errors of the first, second and third estimated PCA coefficients at angle of 90° are shown in Fig. 5, and their mean values are presented in Table 2.

Based on these results, we have the following observations: Firstly, the absence of data augmentation and intensity correction leads to larger errors in estimated PCA coefficients.

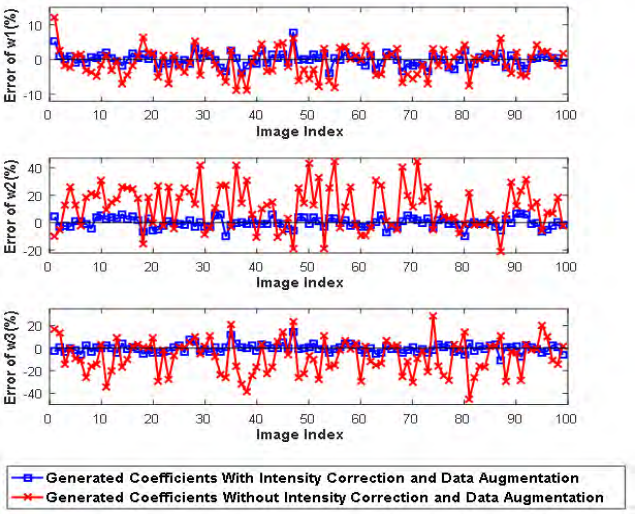

FIGURE 5. The results of PCA coefficients estimation.

TABLE 2. Mean relative error in estimated PCA coefficients.

Case		Mean PCA coefficients errors		
No.	Intensity Correction & Data Augmentation	First	Second	Third
1	Yes	1.1%	2.7%	2.8%
2	No	3.4%	15.3%	13.4%

The errors in results without intensity correction and data augmentation were at least two times larger than the ones with intensity correction and data augmentation. This demonstrated the importance of the data augmentation and intensity correction; Secondly, the error in the first PCA coefficient was significantly less than the errors in other PCA coefficients. This was in line with our expectation, since the first component dominated the lung motion and had the largest effect on the accuracy of volumetric imaging and tumor localization.

2) VOLUMETRIC IMAGES

With the generated PCA coefficients, the deformation vector fields were calculated and the corresponding volumetric images were obtained accordingly. The quantitative comparison between the ground truth volumetric images and the estimated volumetric images were shown in Fig. 6 using the normalized *RMSE* metric. It is evident that the proposed method has much better performance than the MM or MM-FD method on all test subjects.

Table 3 lists the mean value and standard deviation of the normalized RMSE for different cases. Both the MM method and MM-FD method performed relatively well for the scatter-free x-ray projections, with the normalized RMSE below 3%. And MM-FD method had a slightly better accuracy due to the free deformation process. When the projection angle was changed from 0° (acquired from AP direction) to 90° (acquired from LR direction), RMSE increased slightly. However, if the scatter was included in the simulated x-ray

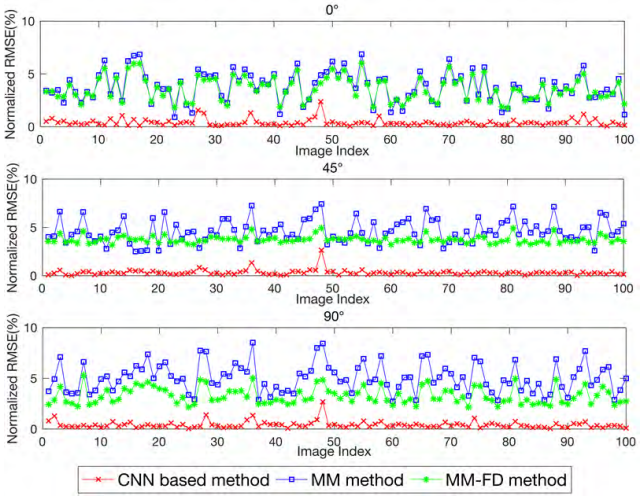

FIGURE 6. The results of normalized RMSE at different angles.

TABLE 3. The statistics of normalized RMSE of different cases.

Case				Statistics of RMSE	
Method	Scatter	Intensity correction	Angle	Mean value	Standard deviation
MM	No	No	0°	2.3%	1.0%
			45°	2.4%	1.1%
			90°	2.9%	1.3%
	Yes	No	0°	4.4%	1.8%
			45°	5.4%	2.3%
			90°	6.0%	3.1%
	Yes	Yes	0°	3.6%	1.3%
			45°	4.4%	1.1%
			90°	5.0%	2.2%
MM-FD	No	No	0°	2.0%	0.8%
			45°	2.1%	0.4%
			90°	2.2%	0.9%
	Yes	No	0°	6.2%	1.9%
			45°	4.2%	1.2%
			90°	4.8%	1.8%
	Yes	Yes	0°	3.5%	1.1%
			45°	3.7%	0.3%
			90°	3.2%	0.8%
CNN based	Yes	Yes	0°	0.3%	0.3%
			45°	0.3%	0.2%
			90°	0.2%	0.3%

projections, the accuracy of MM method and MM-FD method not only suffered dramatically but also declined obviously from 0° to 90° due to the increased level of scatter and noise in the x-ray projections. Applying linear intensity correction could alleviate the effect of scatter to some extent, but the normalized RMSE of these two methods were still more than 3%.

On the contrary, the proposed CNN based method achieved good results even for the projections including scatter. The mean normalized RMSE of the reconstructed volume was below 0.4% in all the cases, much better than the two conventional methods. In addition, there was no obvious difference in accuracy at different angles. These demonstrate the

robustness of the proposed method to the projection angle, scatter and noise of the projections.

3) TUMOR LOCALIZATION

The tumor location was calculated using the resulting DVF from the above experiments. As shown in Fig.7, the result of proposed method was better than the conventional ones. Moreover, we quantitatively analyzed the accuracy of all the methods. Table 4 summarizes the mean value and standard deviation of the errors in tumor localization for different cases. The result is consistent with the result of normalized RMSE for volumetric image comparisons. For MM method, compared with scatter-free projections, the tumor localization accuracy for scatter-included projections declined obviously. Although applying intensity correction could improve the accuracy, there still existed large error in the AP and SI directions. These led to the errors more than 1.4 mm in 3D space. As for MM-FD method, it had an improved accuracy in tumor localization due to the additional registration process. Compared to the results of MM method, the 3D errors at different angles for the scatter-included x-ray projections were reduced by more than 0.1 mm, which meant MM-FD method had better performance in tumor localization than MM method. For the proposed CNN method, the tumor localization errors in three directions were <0.1 mm in all the cases and the errors in 3D were <0.2 mm at different angles. This was clearly much better than the results of the conventional methods. Besides, our method was very robust and showed no performance variation for different angles.

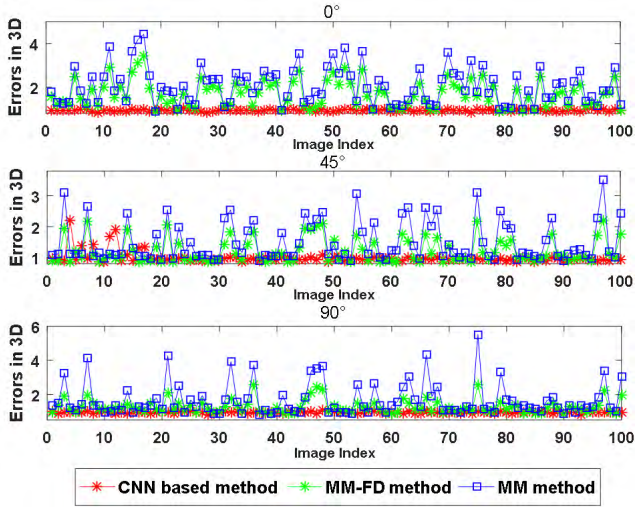

FIGURE 7. The 3D errors in tumor localization for different methods.

Number of training projections and weights of loss function are important parameters involved in the proposed method. In the following experiments, we analyzed their effects on the performance by comparing the results of different settings. As the three evaluation metrics are consistent, we mainly evaluated the mean error of tumor localization in 3D.

TABLE 4. Statistics of errors in tumor localization of different cases.

Method	Scatter	Intensity correction	Angle	Mean error (mm) LR	AP	SI	3D
MM	No	No	0°	0.14	0.76	0.29	0.84
			45°	0.15	0.76	0.34	0.91
			90°	0.18	0.74	0.36	1.02
	Yes	No	0°	0.21	1.90	0.86	2.40
			45°	0.21	1.84	1.58	2.50
			90°	0.28	1.46	1.72	2.38
	Yes	Yes	0°	0.20	1.76	0.48	1.88
			45°	0.18	0.77	1.16	1.58
			90°	0.19	0.73	1.14	1.49
MM-FD	No	No	0°	0.24	0.46	0.23	0.68
			45°	0.28	0.37	0.25	0.59
			90°	0.20	0.38	0.37	0.52
	Yes	No	0°	0.36	1.94	1.24	2.94
			45°	0.34	1.62	1.23	2.35
			90°	0.52	1.34	1.66	2.28
	Yes	Yes	0°	0.22	1.60	0.56	1.73
			45°	0.23	0.74	0.82	1.28
			90°	0.44	0.80	0.76	1.23
CNN based	Yes	Yes	0°	0.06	0.08	0.05	0.12
			45°	0.04	0.07	0.07	0.13
			90°	0.05	0.06	0.06	0.12

B. PARAMETER ANALYSIS

1) NUMBER OF TRAINING PROJECTIONS

The effect of number of training projections is firstly investigated in this section. In our method, the number of the training projections was decided by two processes: the generation of the 3D-CT images from the PCA motion model and the data augmentation for the DRRs. Table 5 shows the results with different parameters in these two processes.

TABLE 5. Results for different number of training projections.

No.	Parameters $N_{3D\text{-}CT}$[a]	Parameters N_{proj}	Mean error of tumor localization with synthetic data at 90°(mm)	Required epochs
1	800	100	0.20	52
2	1000	100	0.12	58
3	1200	100	0.10	70
4	1000	80	0.18	53
5	1000	120	0.11	64

[a] $N_{3D\text{-}CT}$ stands for the number of 3D-CT images generated by sampling the PCA coefficients. N_{proj} infers to the number of training projections generated from single DRR in data augmentation.

According to the results from the case 1 to case 3 in Table 5, it can be seen that we improved the accuracy of tumor localization by utilizing more 3D-CT images generated by sampling the PCA coefficients. However, further increased number of the generated 3D-CT images would lead to training time rising sharply. Actually, when too many 3D-CT images were generated for the training, such as 1200, there would be a significant increase in the number of epochs but a minor

promotion in the final accuracy. Therefore, we obtained 1000 3D-CT images from the PCA motion model to generate the training projections.

Moreover, by comparing the results of case 2 with that of the case 4 and case 5, we can see similar effect on the final accuracy caused by adjusting the number of training projections produced in data augmentation step. Along with the increase of the number of training projections, the final accuracy would be improved, but the training process would also be prolonged. As we can see from the result of case 5, compared to case 2, there was tiny improvement in the accuracy of tumor localization but sharp increase in the training time, which resulted from too many training projections. Hence, to balance the performance and efficiency, we utilized the parameters of case 2.

2) WEIGHTS OF LOSS FUNCTION

Another issue is the weights in the loss function. It is important to set proper weights of PCA coefficients in the loss function to achieve the best performance. Since the first coefficient plays a dominating role in the lung motion, we mainly discuss the effect of the first coefficient weight on our algorithm. To optimize the weight of the first coefficient, we calculated the results from three different cases with synthetic data at 90° and evaluate them with the relative error in PCA coefficients as well as the mean error of tumor localization in 3D. The result is demonstrated in Table 6.

TABLE 6. Results of different weights in loss function.

Case		Results			
No.	Weights of loss function	Error in PCA coefficients (%)			$Error_{Tumor}$ (mm)
		1st	2nd	3rd	
1	$[1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3}]$	1.4	2.3	2.1	0.17
2	$[2/\sqrt{6}, 1/\sqrt{6}, 1/\sqrt{6}]$	1.1	2.7	2.8	0.12
3	$[4/\sqrt{18}, 1/\sqrt{18}, 1/\sqrt{18}]$	1.0	3.1	3.2	0.14

[a] $Error_{Tumor}$ infers to the errors of tumor localization in 3D space with simulated data. from single DRR in data augmentation.

Table 6 shows the second case has the best performance in tumor localization. For case 1, equal weights of the three PCA coefficients would lead to larger error in the first PCA coefficients. Since the first PCA coefficient dominates the respiratory motion, the tumor localization was estimated with low accuracy. The case 3 had the largest weight for the first PCA coefficient, but it also resulted in a decline in the accuracy of localizing tumor position due to the large errors in the estimation of the second and third PCA coefficients. Based on the results above, we use case 2 as the optimal weights.

C. RESULTS FOR REAL PATIENT STUDY

1) TUMOR LOCALIZATION

For each patient, we extracted the tumor contour from the estimated volumetric image and computed the location of the tumor centroid in 3D space. Both the tumor contour and location of tumor centroid in 3D space were projected to the DRR of the reconstructed volumetric image. Fig. 8 shows the clinician-defined tumor contour and the ground truth of tumor centroid position on the CBCT projections at various angles. Moreover, the projection of the estimated tumor contour and calculated tumor centroid location from three methods were also drawn. As it is seen, the estimated tumor location was quite accurate for all three patients at different angles.

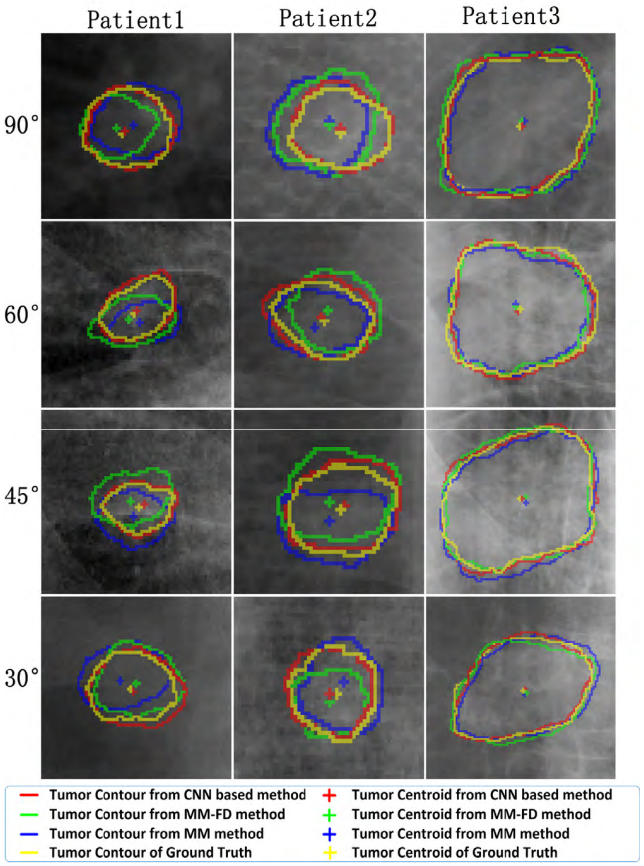

FIGURE 8. The ground truth and result of tumor localization of three methods.

Table 7 lists the tumor localization error for different methods. For the MM method, it had the largest errors in tumor localization, which were more than 1.1 mm in X direction and 1.4 mm in Y direction for all patients at different angles. And the statistics of errors for 3 patients were around 1.5 mm and 1.93 mm in X direction and Y direction, respectively. It also can be seen that when the projection angle varies from 90° to 30°, there was a trend of increasing errors in Y direction, which may be the result from the decreasing visibility of the tumor. Moreover, for MM method, there was an increase in the average errors for single patient when the range of tumor motion becoming large.

As for MM-FD method, it achieved better results compared to MM method. To be more specific, MM-FD method had less mean tumor localization errors in both directions than MM method. However, the minimum errors in X and Y

TABLE 7. Errors of tumor localization for real patients.

Method	Case	Error of tumor localization at different angles (mm)					
		90°	60°	45°	30°	Mean for single patient	Statistics for 3 patients
MM Method	Patient1	(1.63,1.85)	(1.46,2.15)	(1.27,2.09)	(2.12,2.35)	(1.62,2.11)	
	Patient2	(1.57,1.72)	(1.64,1.89)	(1.77,2.06)	(1.79,2.13)	(1.66,1.95)	(1.51 ± 0.21, 1.93 ± 0.19)
	Patient3	(1.36,1.48)	(1.21,1.64)	(1.34,1.89)	(1.18,1.95)	(1.27,1.74)	
MM-FD Method	Patient1	(1.54,1.79)	(1.62,1.95)	(1.88,2.08)	(1.93,2.26)	(1.74,2.02)	
	Patient2	(1.43,1.32)	(1.24,1.58)	(1.65,1.83)	(1.54,1.75)	(1.46,1.62)	(1.46 ± 0.28, 1.67 ± 0.32)
	Patient3	(1.08,1.24)	(1.13,1.27)	(1.29,1.48)	(1.21,1.55)	(1.19,1.38)	
CNN based method	Patient1	(0.63,0.74)	(0.76,0.65)	(0.74,0.86)	(0.69,0.89)	(0.71,0.79)	
	Patient2	(0.23,0.83)	(0.88,0.78)	(0.72,0.69)	(0.82,0.54)	(0.66,0.71)	(0.71 ± 0.05, 0.76 ± 0.04)
	Patient3	(0.74,0.69)	(0.77,0.82)	(0.75,0.79)	(0.85,0.82)	(0.77,0.78)	

directions for all cases were still more than 1.0 mm and 1.2 mm, respectively. Besides, similar to MM method, when the projection angle changed from 90° to 30° or the range of tumor motion increased, there was an increase in the errors of tumor localization for MM-FD method.

For the CNN based method, the maximum value of tumor localization errors in two directions were both <0.9 mm; the average errors for single patient were <0.8 mm in two directions; and the mean values of the errors for all the three patients were only 0.71 mm and 0.76 mm in X and Y direction. Moreover, there was no obvious difference in the performance of the CNN based method for all the three patients with different ranges of tumor motion and projection angles. These results demonstrate that the proposed method was more accurate and robust.

2) HEIGHTS OF DIAPHRAGMS

For the CBCT projections where the tumor was basically not visible, we can still estimate the position and shape of the tumor contour in 3D space. As shown in Fig.9, the estimated tumor contours were drawn on DRRs and CBCT projections, and the heights of the diaphragm in DRRs were highly consistent with that in CBCT projections.

To evaluate the efficacy of the CNN based method quantitatively, we compared the diaphragm height estimation results of the proposed CNN based method with that from the two conventional methods. Similar with the result of the tumor localization, MM method had the largest error in estimating the position of the diaphragm, whose mean value was 1.6 mm for all the cases. MM-FD method performed better than MM method due to the free-deformation registration process. However, it still had an average error of 1.2 mm because of the large inconsistency between CBCT projections at 0° and the corresponding DRRs. As for the CNN based method, the results of the first two patients were consistent with the ground truth, and error of third patient was only 0.8 mm, which meant, compared to the conventional method, the proposed method had higher accuracy in estimating the position of the diaphragm for the projections at 0°.

D. COMPUTATION TIME

In this work, we ran the experiments on a PC with Intel Core i7-5960X (3.0 GHz), 32GB memory and a NVIDIA GTX 1080 Ti graphic card. For the proposed method, it would cost about 30-40 minutes for the network training. And during the application stage, it took 36 milliseconds to apply the CNN model and calculate the tumor position from single x-ray projection, which allows for real-time tumor tracking during treatment. If the corresponding information of patient anatomy is needed, our method is able to reconstruct the volumetric image within 1 second, which can be further accelerated by GPU.

IV. DISCUSSION AND CONCLUSION

In this study, we developed a CNN-based regression method for tumor localization from single x-ray projection. A motion model was build using PCA with prior knowledge for the estimation of deformation vector fields. Unlike the previous methods, we utilized a CNN model to recover the ill-conditioned nonlinear mapping from single x-ray projection to tumor motion. Moreover, novel intensity correction process and data augmentation step were proposed to handle intensity discrepancy between DRR and CBCT projection caused by the scatter and quantum noise. The proposed method was validated using simulation and actual patient images, and experimental results showed that our algorithm could localize the tumor position and reconstruct the volumetric images with significantly higher accuracy than the conventional registration-based methods. Furthermore, our algorithm was capable of real-time tumor tracking.

In addition, we found that, for both MM and MM-FD methods, inappropriate initial values of the PCA coefficients for registration could lead to decreased accuracy in tumor localization. Since the first PCA coefficient usually has a large range, it is difficult to set an appropriate initial value to achieve the best performance of the algorithms for all the x-ray projections. On the contrary, our approach had no parameter that need to be tuned for different x-ray projections. Once the CNN regression model was trained, the performance of our approach had no preference for different x-ray projections. Therefore, compared with the MM method and MM-FD method, the performance of our method was more consistent in regard to projection angles.

In our current method, a CBCT scan and its associated raw projection images of the patient are needed for the intensity correction. This may seem to be an obstacle in the practical

implementation, as the data processing and CNN training take some time and may not be completed while the patient is on the treatment table. In actual practice, the CBCT and projection images are often acquired during the pre-treatment simulation procedure which occurs typically one day before the first fraction treatment, so patient positioning and images can be verified and approved and marks can be placed on patient. This should provide ample time for the CNN training. During the actual treatment, the application of the CNN model is instantaneous. In addition, several improvements can be made in the procedure to speed up the processing to make it nearly real-time for situations that the pre-treatment simulation is not available. Firstly, current implementation utilized hard disk for training data sharing (about 60 GB) between different procedures. Saving all the training data in memory could greatly reduce the time spent. Secondly, using more powerful GPUs to train the CNN model could also speed-up the training process. In addition, the proposed intensity correction procedure does not need the accurate information of the CBCT scan from the same patient, but the discrepancies between the 4D-CT and 3D-CBCT, therefore, the correction can be based on the information from a patient-population based average. However, further investigation is needed to validate this.

For the future work, since the proposed method could only be used for IMRT, we will develop a method that could localize the real time tumor position from CBCT projections at varied angle, which will be preferable for VMAT.

REFERENCES

[1] V. Vishnevskiy, T. Gass, G. Szekely, C. Tanner, and O. Goksel, "Isotropic total variation regularization of displacements in parametric image registration," *IEEE Trans. Med. Imag.*, vol. 36, no. 2, pp. 385–395, Feb. 2017. doi: 10.1109/TMI.2016.2610583.

[2] X. Li, Z. Deng, Q. Deng, L. Zhang, T. Niu, and Y. Kuang, "A novel deep learning framework for internal gross target volume definition from 4D computed tomography of lung cancer patients," *IEEE Access*, vol. 6, pp. 37775–37783, 2018. doi: 10.1109/ACCESS.2018.2851027.

[3] C.-C. Shieh, P. J. Keall, Z. Kuncic, C.-Y. Huang, and I. Feain, "Markerless tumor tracking using short kilovoltage imaging arcs for lung image-guided radiotherapy," *Phys. Med. Biol.*, vol. 60, no. 24, pp. 9437–9454, 2015. doi: 10.1088/0031-9155/60/24/9437.

[4] T. Lin, L. I. Cervino, X. Tang, N. Vasconcelos, and S. B. Jiang, "Fluoroscopic tumor tracking for image-guided lung cancer radiotherapy," *Phys. Med. Biol.*, vol. 54, no. 4, pp. 981–992, 2009. doi: 10.1088/0031-9155/54/4/011.

[5] B. D. de Senneville, A. El Hamidi, and C. Moonen, "A direct PCA-based approach for real-time description of physiological organ deformations," *IEEE Trans. Med. Imag.*, vol. 34, no. 4, pp. 974–982, Apr. 2015. doi: 10.1109/TMI.2014.2371995.

[6] S. Xie, C. Yang, Z. Zhang, and H. Li, "Scatter artifacts removal using learning-based method for CBCT in IGRT system," *IEEE Access*, vol. 6, pp. 78031–78037, 2018. doi: 10.1109/ACCESS.2018.2884704.

[7] Q. Zhang *et al.*, "A patient-specific respiratory model of anatomical motion for radiation treatment planning," *Med. Phys.*, vol. 34, no. 12, pp. 4772–4781, 2007. doi: 10.1118/1.2804576.

[8] J. Wasza *et al.*, "Real-time respiratory motion analysis using 4-D shape priors," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 3, pp. 485–495, Mar. 2016. doi: 10.1109/TBME.2015.2463769.

[9] Y. Xu *et al.*, "A method for volumetric imaging in radiotherapy using single X-ray projection," *Med. Phys.*, vol. 42, no. 5, pp. 2498–2509, 2015. doi: 10.1118/1.4918577.

[10] R. Li *et al.*, "3D tumor localization through real-time volumetric X-ray imaging for lung cancer radiotherapy," *Med. Phys.*, vol. 38, no. 5, pp. 2783–2794, 2011. doi: 10.1118/1.3582693.

[11] Y. Zhang, F.-F. Yin, W. P. Segars, and L. Ren, "A technique for estimating 4D-CBCT using prior knowledge and limited-angle projections," *Med. Phys.*, vol. 40, no. 12, pp. 121701–121710, 2013. doi: 10.1118/1.4825097.

[12] W. Yang, H. Zhang, J. Yang, J. Wu, X. Yin, and Y. Chen, "Improving low-dose CT image using residual convolutional network," *IEEE Access*, vol. 5, pp. 24698–24705, 2017. doi: 10.1109/ACCESS.2017.2766438.

[13] X. Zhang, N. Homma, K. Ichiji, Y. Takai, and M. Yoshizawa, "Tracking tumor boundary in MV-EPID images without implanted markers: A feasibility study," *Med. Phys.*, vol. 42, no. 5, pp. 2510–2523, 2015. doi: 10.1118/1.4918578.

[14] S. Christodoulidis, M. Anthimopoulos, L. Ebner, A. Christe, and S. Mougiakakou, "Multisource transfer learning with convolutional neural networks for lung pattern analysis," *IEEE J. Biomed Health Inform.*, vol. 21, no. 1, pp. 76–84, Jan. 2017. doi: 10.1109/JBHI.2016.2636929.

[15] W. P. Segars, G. Sturgeon, S. Mendonca, J. Grimes, and B. M. W. Tsui, "4D XCAT phantom for multimodality imaging research," *Med. Phys.*, vol. 37, pp. 4902–4915, Sep. 2010. doi: 10.1118/1.3480985.

[16] M. S. Brown, M. F. McNitt-Gray, J. G. Goldin, R. D. Suh, J. W. Sayre, and D. R. Aberle, "Patient-specific models for lung nodule detection and surveillance in CT images," *IEEE Trans. Med. Imag.*, vol. 20, no. 12, pp. 1242–1250, Dec. 2001. doi: 10.1109/42.974919.

[17] R. Wang, X. Liang, X. Zhu, and Y. Xie, "A feasibility of respiration prediction based on deep Bi-LSTM for real-time tumor tracking," *IEEE Access*, vol. 6, pp. 51262–51268, 2018. doi: 10.1109/ACCESS.2018.2869780.

[18] M. Chen, K. Cao, Y. Zheng, and R. A. C. Siochi, "Motion-compensated mega-voltage cone beam CT using the deformation derived directly from 2D projection images," *IEEE Trans. Med. Imag.*, vol. 32, no. 8, pp. 1365–1375, 2013. doi: 10.1109/TMI.2012.2231694.

[19] W. Cong, Y. Xi, and G. Wang, "X-ray fluorescence computed tomography with polycapillary focusing," *IEEE Access*, vol. 2, pp. 1138–1142, 2014. doi: 10.1109/ACCESS.2014.2359831.

[20] T. Lin, R. Li, X. Tang, J. G. Dy, and S. B. Jiang, "Markerless gating for lung cancer radiotherapy based on machine learning techniques," *Phys. Med. Biol.*, vol. 16, no. 2, pp. 267–269, 2009. doi: 10.1088/0031-9155/54/6/010.

[21] D. Sarrut, B. Delhay, P.-F. Villard, V. Boldea, M. Beuve, and P. Clarysse, "A comparison framework for breathing motion estimation methods from 4-D imaging," *IEEE Trans. Med. Imag.*, vol. 26, no. 12, pp. 1236–1248, Dec. 2007.

[22] J. Ehrhardt, R. Werner, A. Schmidt-Richberg, and H. Handels, "Statistical modeling of 4D respiratory lung motion using diffeomorphic image registration," *IEEE Trans. Med. Imag.*, vol. 30, no. 2, pp. 251–265, Feb. 2011. doi: 10.1109/TMI.2010.2076299.

[23] D. van Eeden and F. du Plessis, "EGS_cbct: Simulation of a fan beam CT and RMI phantom for measured HU verification," *Med. Phys.*, vol. 32, no. 10, pp. 1375–1380, 2016.

RAN WEI received the bachelor's degree in control engineering from Beihang University, Beijing, in 2011, where he is currently pursuing the Ph.D. degree in pattern recognition. His research interests include the medical image processing and deep learning.

FUGEN ZHOU received the Ph.D. degree in pattern recognition from Beihang University, Beijing, China, in 2006, where he is currently a Professor with the Image Processing Center. His research interests include the medical image processing techniques, image registration and fusion techniques, digital video processing techniques, and deep learning techniques.

BO LIU received the Ph.D. degree from Beihang University, Beijing, in 2015, where he is currently an Assistant Professor with the Image Processing Center. His research interests include biomedical image processing, pattern recognition, and medical physics.

XIANGZHI BAI received the B.S. and Ph.D. degrees from Beihang University, Beijing, in 2003 and 2009, respectively, where he is currently a Professor with the Image Processing Center. He holds 20 national invention patents. He has published over 100 international journals and conference papers in the fields of mathematical morphology, fuzzy clustering, image analysis, pattern recognition, and bioinformatics. He also acts as an active Reviewer of around 40 international journals and conferences.

DONGSHAN FU received the Ph.D. degree in electrical engineering from the University of Rochester, New York, in 2000. He is currently a Senior Researcher with the Beijing Advanced Innovation Center for Biomedical Engineering, Beihang University. His research interests include multiple modality image-guided positioning and tracking, AI treatment planning, intelligence robotic radiosurgery, and cardiac radiosurgery.

YONGBAO LI received the Ph.D. degree in nuclear science and technology from Tsinghua University, Beijing, in 2015. He is currently a Medical Physicist and an Assistant Professor with the Department of Radiation Oncology, Sun Yat-sen University Cancer Center. His research interest includes the radiation therapy physics, especially on treatment plan optimization and dose calculation for intensity modulated radiotherapy.

BIN LIANG received the Ph.D. degree from Beihang University, Beijing, in 2017. He is currently a Physicist with the National Cancer Center/Hospital, Chinese Academy of Medical Sciences. His research interests include biomedical image processing, pattern recognition, and medical physics.

QIUWEN WU received the Ph.D. degree in physics from Columbia University, New York, in 1994. He is currently a Professor with the Department of Radiation Oncology, Duke University Medical Center. His research interests include intensity-modulated radiation therapy (IMRT), volumetric modulated arc therapy (VMAT), electron arc therapy (EAT), and image-guided radiation therapy (IGRT).