Simulation of Random Deformable Motion in Soft-Tissue Cone-Beam CT with Learned Models

Y. Hu[a], H. Huang[b], J. H. Siewerdsen[a, b], W. Zbijewski[b], M. Unberath[a],
C. R. Weiss[c], and A. Sisniega [*b]
[a]Dept. of Computer Science, Johns Hopkins University, Baltimore, MD, USA
[b]Dept of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, USA
[c]Russell H. Morgan Department of Radiology, Johns Hopkins University, Baltimore, MD, USA

ABSTRACT

Cone-beam CT (CBCT) is widely used for guidance in interventional radiology but it is susceptible to motion artifacts. Motion in interventional CBCT features a complex combination of diverse sources including quasi-periodic, consistent motion patterns such as respiratory motion, and aperiodic, quasi-random, motion such as peristalsis. Recent developments in image-based motion compensation methods include approaches that combine autofocus techniques with deep learning models for extraction of image features pertinent to CBCT motion. Training of such deep autofocus models requires the generation of large amounts of realistic, motion-corrupted CBCT. Previous works on motion simulation were mostly focused on quasi-periodic motion patterns, and reliable simulation of complex combined motion with quasi-random components remains an unaddressed challenge.

This work presents a framework aimed at synthesis of realistic motion trajectories for simulation of deformable motion in soft-tissue CBCT. The approach leveraged the capability of conditional generative adversarial network (GAN) models to learn the complex underlying motion present in unlabeled, motion-corrupted, CBCT volumes. The approach is designed for training with unpaired clinical CBCT in an unsupervised fashion. This work presents a first feasibility study, in which the model was trained with simulated data featuring known motion, providing a controlled scenario for validation of the proposed approach prior to extension to clinical data. Our proof-of-concept study illustrated the potential of the model to generate realistic, variable simulation of CBCT deformable motion fields, consistent with three trends underlying the designed training data: i) the synthetic motion induced only diffeomorphic deformations – with Jacobian Determinant larger than zero; ii) the synthetic motion showed median displacement of 0.5 mm in regions predominantly static in the training (e.g., the posterior aspect of the patient laying supine), compared to a median displacement of 3.8 mm in regions more prone to motion in the training; and iii) the synthetic motion exhibited predominant directionality consistent with the training set, resulting in larger motion in the superior-inferior direction (median and maximum amplitude of 4.58 mm and 20 mm, > 2x larger than the two remaining direction). Together, the proposed framework shows the feasibility for realistic motion simulation and synthesis of variable CBCT data.

Keywords: Interventional CBCT, Motion Simulation, Motion Compensation, Deep Learning.

1. INTRODUCTION

Cone-beam CT (CBCT) is becoming widespread for guidance and intraprocedural imaging in interventional radiology, but it suffers from relatively long image acquisition time that makes it prone to degradation from patient motion. Motion in interventional CBCT displays a complex nature and a wide variety, spanning from rigid aperiodic motion (as in brain CBCT) to multi-source deformable motion in abdominal imaging, mixing quasi-periodic motion components (e.g., respiratory) with aperiodic, quasi-random motion (e.g., peristalsis).

Motion compensation for interventional CBCT has gained significant attention, with image-based approaches including autofocus methods based on handcrafted metrics [1-3], and methods leveraging deep convolutional neural networks (CNNs) to directly learn motion trajectories from distortion patterns [4], or to learn features associated to motion effects that are aggregated into deep autofocus metrics [5, 6]. Common to those approaches is the need for simulation methods that allow the generation of large amounts of realistic, motion-corrupted, CBCT data to enable training and evaluation. The fidelity of simulated datasets to experimental CBCT data is of dual nature: i) the data should show a realistic image

*asisniega@jhu.edu; phone 1 443-285-1328

7th International Conference on Image Formation in X-Ray Computed Tomography,
edited by Joseph Webster Stayman, Proc. of SPIE Vol. 12304, 1230413
 · 0277-786X · doi: 10.1117/12.2646720

appearance, attenuation pertinent to CBCT, and realistic noise and artifacts patterns; and, ii) the synthetic motion should be true to motion observed in clinical CBCT. Recent work showed the capability of fulfilling the first condition via accurate models of the CBCT imaging chain and biological tissues [7]. However, the generation of realistic motion patterns remains an open question in interventional CBCT.

Previous efforts to motion simulation yielded highly accurate models of temporal motion patterns and tissue deformation for quasi-periodic (respiratory and cardiac) motion simulation [8]. However, those models did not provide mechanisms to integrate the remaining sources of motion present in interventional CBCT, some of which feature a highly unpredictable nature (e.g., head involuntary motion or peristaltic motion).

Recent advances in deep learning-based data synthesis architectures and conditional generative adversarial network (GAN) models, have shown the capability of such approaches to learn features associated with complex underlying characteristics of the training data that, when combined with random perturbation models, allowed the synthesis of highly realistic, variable, datasets. Such methods were recently proposed for simulation of non-periodic respiratory motion in 4D CT data synthesis for image-guided radiotherapy applications [9].

In this work we hypothesize that conditional GAN models can be used to learn the underlying motion characteristics in unpaired, motion-corrupted, clinical datasets, with no prior knowledge or prior assumptions on motion nature. A GAN model is proposed, and a proof-of-concept study is presented. This proof-of-concept study used simulated data with known motion fields. The training was completely agnostic to the known motion, analogous to training with clinical datasets, but knowledge of the true motion pattern allowed validation of the characteristics of the random synthetic motion generated by the trained model.

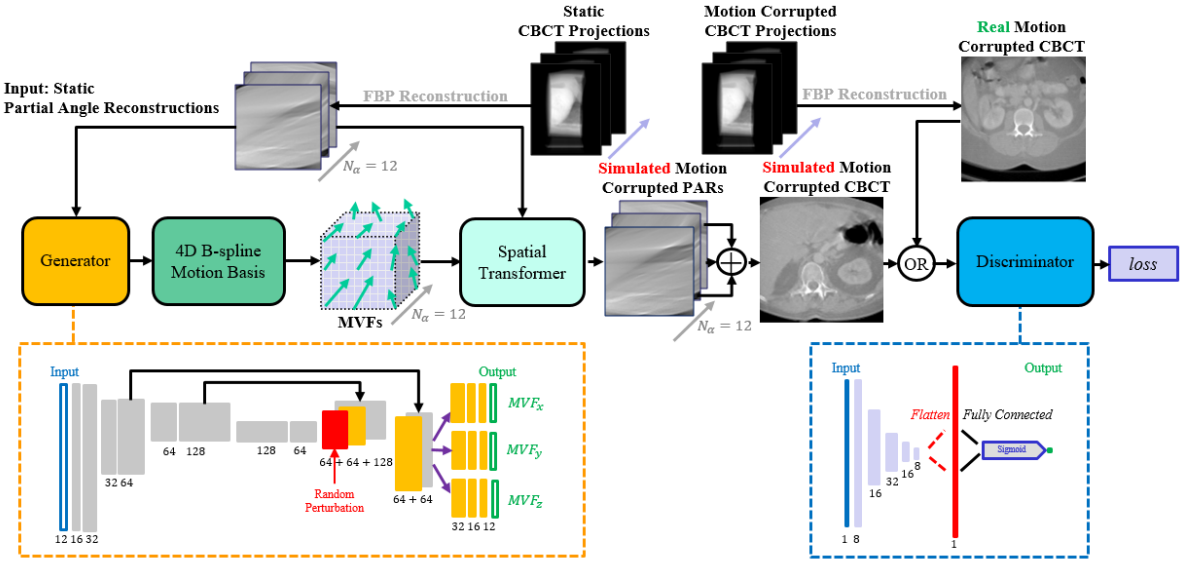

Figure 1. Schematic depiction of the realistic motion simulation framework, based on a deep generative adversarial network model. Motion-free simulated CBCT projections are divided into $N_\alpha = 12$ groups of consecutive projections that generate a set of N_α. The generator network, based on the U-NET architecture, receives at input the set of static PARs and a random perturbation of the latent space. The sparse 4D MVF output by the generator induces the deformation into the individual PARs that were added together into the final motion-corrupted volume. For GAN training, a discriminator receives as input the generated simulated motion volumes and samples of the real motion-corrupted volumes from the training dataset.

2. MATERIAL AND METHODS

2.1 Learning Complex Deformable Motion with a GAN model

The proposed GAN architecture is illustrated in Fig. 1. The proposed approach leveraged the concept of Partial Angle Reconstruction (PAR) combined with spatial transformer modules, previously used in CBCT motion compensation [4]. A complete, motion-free, CBCT projection dataset, with a total of 360 projections, is reconstructed into N_α = 12 PAR volumes of 512 × 512 × 128 voxels, each containing the backprojection from 30 consecutive projection views. Those PARs are the input to the generator network, which outputs N_α sets of 36 × 36 × 12 B-spline coefficients that serve as a lower dimensionality representation for each of the N_α motion vector fields (MVFs) representing the simulated 4D

deformable motion of the volume. Dense MVFs, with size equal to that of the PAR volumes, are then generated via B-spline interpolation. The set of dense MVFs and the original motion-free PARs are then input to a spatial transformer module that applies the simulated deformation to each of the PARs and add them together to obtain a final motion-corrupted volume. In the resulting architecture PARs are considered static, effectively assuming a piecewise constant temporal motion trajectory. A discriminator network was used to provide a GAN loss discriminating between real and simulated motion-corrupted volumes. The architecture of each of the components is discussed below.

Generator: The generator featured a non-symmetrical 3D UNet-like [10] structure with a 3-stage encoding branch, and 2-stage decoding branch. The encoding branch received the motion-free PARs as the input and extracted into the latent space features associated to structural content of the image associated with motion characteristics. This way, the input, motion-free PARs act as the condition variable of the conditional GAN architecture. Each stage on the encoder branch included a set of two 3×3×3 convolution layers, batch normalization, leaky ReLU activation, and a final 2x max pooling layer.

The set of latent space features were combined with a Gaussian random perturbation field to generate random, distinct, motion patterns for a given input condition during both training and inference time. The set of latent features and random perturbation entered the decoder branch, with 2 stages implementing a 3×3×3 convolution layer, a batch normalization layer, and leaky ReLU activation, followed by a 3×3×3 transposed convolution for up-sampling of the feature maps. Skip connections were placed between equivalent levels of the encoder and decoder branches. The output of the decoder is input to three branches implementing a cascade of two 3×3×3 convolution layers, with leaky ReLU activations, that generate the B-spline coefficients for the directional components of the MVFs in the antero-posterior (AP), lateral (LAT), and superior-inferior (SI) directions.

Discriminator: The discriminator acts on motion-corrupted CBCT volumes to predict whether the input comes from a real or simulated instance. During training, the Binary Cross Entropy (BCE) loss was calculated for the simulated and real datasets, and the total loss was defined as the average of both.

In the proposed model, the discriminator featured a cascade of 5 convolution layers (4×4×4 kernel), followed by batch normalization, leaky ReLU activation, and a dropout layer (0.2 dropout). The final fully connected layer (with sigmoid activation) acted on the flattened set of features.

2.2 Data Generation and Motion Model

For this proof-of-concept study, training and validation data were generated from 70 Multi-detector CT (MDCT) abdominal datasets from the TCIA Lymph Node Abdomen collection. 60 distinct MDCT instances were used for training, 5 for validation, and 5 for testing. For each source MDCT volume, we randomly selected a subvolume of 128 mm length at a random longitudinal position within the abdomen. The subvolume was then forward projected using a high-fidelity CBCT model with a geometry with source-to-detector distance of 1200 mm, and source-to-axis distance of 785 mm. The detector was modeled as a flat panel with 576 × 440 pixels (0.616 mm isotropic pixel size).

Motion corrupted datasets were obtained by inducing deformable motion during forward projection. The simulated motion field followed a cosine temporal trajectory with random frequency between 0.75 – 1.25 cycles per scan and random phase. Spatial distribution of motion amplitude was modelled as an elliptical field with maximum amplitude (randomly set between 10 and 25 mm) at the center, and randomly placed at a soft-tissue region of the volume. The amplitude faded following a Gaussian decay curve that reached zero at the ellipse axis length, randomly chosen from 200 to 300 mm in the medial-lateral (LAT) direction and between 100 and 150 mm in the antero-posterior (AP) direction. Motion amplitude was kept constant across slices. Motion direction was randomly chosen, allocating between 60% and 80% of motion to the SI direction and the rest to the AP direction, with no lateral motion.

To avoid unrealistic large motion of the spine region, the center of the spine was detected in the volume and a cylindrical motion-exclusion mask with 100 mm radius was defined. The mask performed a smooth transition from one to zero and multiplied the motion field, to minimize the motion in the spine, as illustrated in Fig 2.

The motion-corrupted datasets featured 3 distinct properties that were used for validation of the GAN capability for inference of consistent motion instances: i) the induced motion was composed of diffeomorphic deformations; ii) the spine region remained nearly static for all training instances; and, iii) the majority of the motion was allocated to the SI direction with the rest in the AP direction.

Motion-corrupted datasets were reconstructed into volumetric grids of $512 \times 512 \times 128$ voxels with $0.5 \times 0.5 \times 0.5$ mm^3 voxel size, and motion-free cases were reconstructed into 12 PARs with equivalent parameters. The PARs were downsampled to $128 \times 128 \times 32$ voxels for input to the generator.

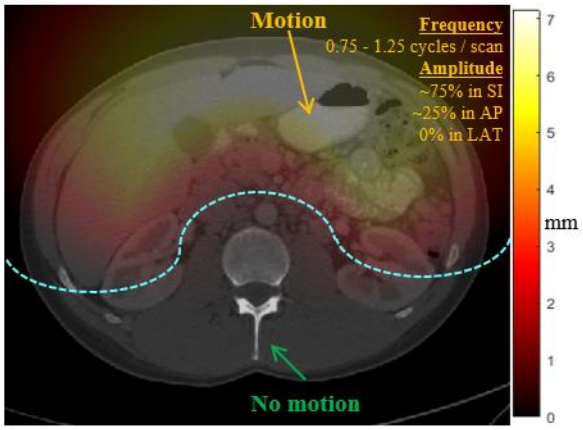

Figure 2. Example training case. Deformable motion was induced following with a motion field with randomly selected central location and amplitude. In this proof-of-concept study, the spinal region was automatically detected and an exclusion mask with smooth boundaries was defined (see cyan delineation). The motion field was attenuated to preserve the spine region quasi-static.

2.3 Network training and validation studies

A total of 720 motion-corrupted volumes based on 60 anatomical instances were included in the training set, and 15 instances from 5 separate anatomical structures were used for validation. Each dataset contained a motion-free collection and a motion-corrupted collection. During training, a sample with N_α = 12 PARs from the motion-free collection was randomly selected as input to the generator, while one sample from motion-corrupted collection was input to the discriminator. Training was performed with an unbalanced scheme in which the generator is updated every 1 batch while the discriminator is updated every 2 batches for 100 epochs, with a batch size of 12. We used the ADAM optimizer for both generator and discriminator with learning rates of 10^{-5} and 10^{-4}, respectively. BCE Loss was selected as objective function to be maximized by generator while minimized by discriminator.

To validate the results, we used a test set of 15 samples based on 5 anatomical structures not seen by the network. For testing, static PARs were input to the generator together with the Gaussian random perturbation. Validity of the generated motion fields was validated via measurements of diffeomorphism, based on the determinants of the Jacobian of the deformation, and metrics of average displacement at regions of maximum motion and regions static in the training set. Furthermore, directional components of the synthetic motion were evaluated in comparison with underlying trends in the training dataset.

3. RESULTS

Figure 3 illustrates the predominant soft-tissue nature of the simulated motion, as well as its diffeomorphic nature. An example simulated MVF is shown if Fig. 3A, demonstrating the majority of the deformation induced to anterior soft-tissue regions with minimal deformation towards the central posterior area, where the spine is located. Fig. 3B shows the accumulated distribution of Jacobian determinant values across the ensemble of test datasets. The induced deformable motion vector fields consistently show Jacobian determinant values larger than zero, consistent with diffeomorphic motion.

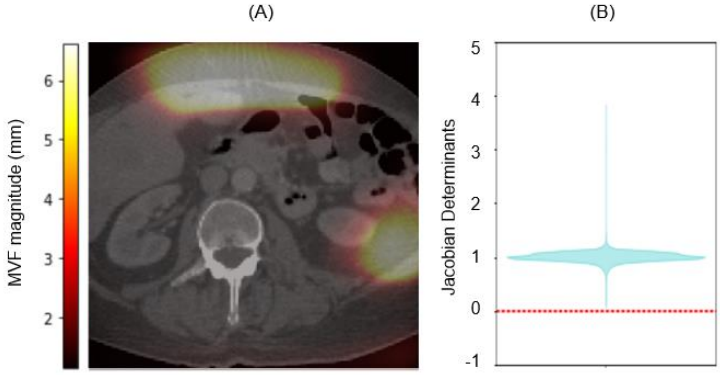

Figure 3. (A) Motion vector field on an example inference instance, with motion predominantly present in soft-tissue structures. (B) Validation of the diffeomorphism of the generated deformable motion fields for the aggregated test set.

Fig. 4A shows the average displacement for the aggregated motion synthesis dataset, obtained by adding the absolute value of the motion amplitude for each time point (viz. PAR) and normalizing the result by the total number of PARs (N_α = 12). Average displacement was evaluated in a soft-tissue region in the anterior area of the abdomen and in a region inside the spine. Results show displacement values in anterior soft-tissue areas of 3.9 ± 2.5 mm, while spine regions showed minimal motion, with average displacement of 0.5 ± 0.0 mm. Fig. 4B shows the directional properties of the random motion instances synthesized by the GAN model. Consistent with the trends underlying the training data, the synthetic motion fields exhibited larger motion in the SI direction with median amplitude of 4.58 mm and ranging upwards of 20 mm consistent with voluntary or involuntary respiratory motion.

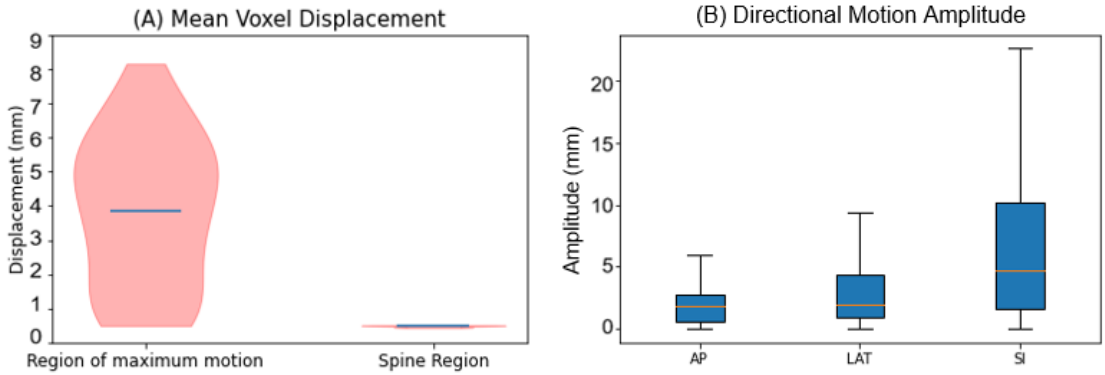

Figure 4. Quantitative evaluation of the synthetic random motion. (A) Average displacement of voxels in the maximum amplitude and in the spine regions, showing preservation of the quasi-static nature of the spine. (B) Motion amplitude in the AP, LAT, and SI directions for the generated random motion vector fields.

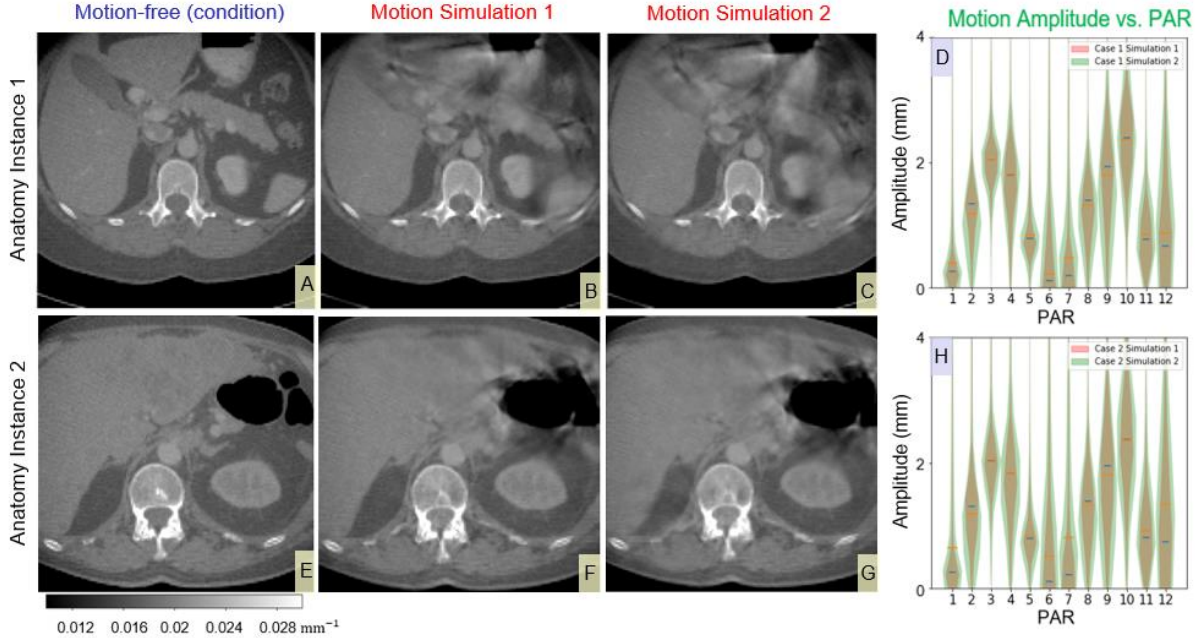

Figure 5. Example motion simulation cases for two motion-free source anatomical instances (A, E), and two instances of the latent space random perturbation (B, C, and F, G), showing distinct motion artifacts but respecting the learned properties in terms of motion distribution and main direction. (D, H) Distribution of motion amplitude as a function of time (viz. PAR index) for the random realization in (B, F) in red and the motion realization in (C, G) in green.

Motion amplitude in the AP and LAT directions was lower, with median values of 2.02 mm, and 2.18 mm, respectively. The comparable amplitude observed for the AP and LAT directions illustrate the challenge in differentiation of motion patterns that can result in similar artifacts, as lateral motion was minimal in the training data.

Note that while the current design did not implement any control mechanism on the output motion amplitude, several options for such controlled simulation can be easily integrated, including coarse stratification of the training data into mild,

moderate, or severe cases within a semi-supervised training strategy; normalization of the output motion fields; or controlled scaling of the latent space random perturbation.

Validation of the realism of the synthetic motion-corrupted datasets and of capability to generate distinct motion for a given input is shown in Fig. 5. Image results in Fig. 5 show distinct, realistic motion artifacts in soft-tissue regions, with minimal distortion of the (static) spine. Quantitative evaluation of motion amplitude in Fig. 5D and Fig. 5H illustrates the generation of variable motion patterns for single input conditions.

4. DISCUSSION AND CONCLUSION

This work presented an adversarial model for simulation of realistic, random, deformable motion in CBCT using motion-corrupted datasets with no prior assumptions on the motion characteristics. The framework was evaluated in a controlled study in which the properties of the random synthetic motion fields were compared with known motion trends underlying in the training data cohort. The model was able to generate distinct motion instances, while replicating principal properties of the training dataset, such as diffeomorphism, proper spatial distribution of motion amplitude (maximized anteriorly and minimized posteriorly), and predominantly SI motion direction in agreement with learned patterns. The results enable generation of large training datasets for development of deep learning autofocus methods.

ACKNOWLEDGEMENTS

This work was supported by the National Institute of Health under Grant R01-EB-030547.

REFERENCES

[1] Sisniega, A., Stayman, J. W., Yorkston, J., Siewerdsen, J. H., & Zbijewski, W. "Motion compensation in extremity cone-beam CT using a penalized image sharpness criterion." *Physics in medicine and biology* vol. 62,9: 3712-3734 (2017).

[2] Hahn, J., Bruder, H., Rohkohl, C., Allmendinger, T., Stierstorfer, K., Flohr, T., & Kachelrieß, M. "Motion compensation in the region of the coronary arteries based on partial angle reconstructions from short-scan CT data." *Medical Physics* 44: 5795–5813 (2017).

[3] Capostagno, S., Sisniega, A., Stayman, J. W., Ehtiati, T., Weiss, C. R., & Siewerdsen, J. H. "Deformable motion compensation for interventional cone-beam CT." *Physics in medicine and biology* vol. 66,5 055010. 17 Feb. (2021).

[4] Maier, J., Lebedev, S., Erath, J., Eulig, E., Sawall, S., Fournié, E., Stierstorfer, K., Lell, M., & Kachelrieß, M. "Deep learning-based coronary artery motion estimation and compensation for short-scan cardiac CT." *Medical Physics* 48: 3559-3571(2021).

[5] Preuhs, A. Manhart, M., Roser, P., Hoppe, E., Huang, Y., Psychogios, M., Kowarschik, M., & Maier, A. "Appearance Learning for Image-Based Motion Estimation in Tomography," *IEEE Trans. Med. Imaging*, vol. 39, no. 11, pp. 3667–3678, Nov. (2020).

[6] Huang, H., Siewerdsen, J. H., Zbijewski, W., Weiss, C. R., Unberath, M., Ehtiati, T., & Sisniega, A. "Reference-free learning-based similarity metric for motion compensation in cone-beam CT." *Physics in medicine and biology* vol 67, 12 10.1088/1361-6560/ac749a. 16 Jun. (2021)

[7] Wu, P., Sisniega, A., Uneri, A., Han, R., Jones, C. K., Vagdragi, P., Zhang, X., Luciano, M., Anderson, W. S., & Siewerdsen, J. H. "Using Uncertainty in Deep Learning Reconstruction for Cone-Beam CT of the Brain." *16th Virtual International Meeting on Fully 3D Image Reconstruction in Radiology and Nuclear Medicine*. (2021)

[8] Segars, W. P., Mahesh, M., Beck, T. J., Frey, E. C., & Tsui, B. M. "Realistic CT simulation using the 4D XCAT phantom." *Medical physics* vol. 35,8: 3800-8 (2008).

[9] Chang, Y., Jiang, Z., Segars, W. P., Zhang, Z., Lafata, K., Cai, J., Yin, F. F., & Ren, L. "A generative adversarial network (GAN)-based technique for synthesizing realistic respiratory motion in the extended cardiac-torso (XCAT) phantoms." *Physics in medicine and biology* vol. 66,11 10.1088/1361-6560/ac01b4. 31 May. (2021).

[10] Ronneberger, O., Fischer, P., & Brox, T. "U-Net: Convolutional Networks for Biomedical Image Segmentation." *Medical Image Computing and Computer-Assisted Intervention*. Lecture Notes in Computer Science(), vol 9351. Springer, Cham (2015)