Projection-based CBCT Motion Correction using Convolutional LSTMs

Joscha Maier[1], Timothy Herbst[1,2], Stefan Sawall[1,2], Marcel Arheit[3], Pascal Paysan[3], and Marc Kachelrieß[1,2]

[1]Division of X-Ray Imaging and Computed Tomography, German Cancer Research Center (DKFZ), Heidelberg, Germany
[2]Heidelberg University, Heidelberg, Germany
[3]Varian Medical Systems Imaging Lab, GmbH, Baden-Dättwil, Switzerland

GERMAN CANCER RESEARCH CENTER IN THE HELMHOLTZ ASSOCIATION

Research for a Life without Cancer

Purpose

Several cone-beam CT (CBCT) application rely on time-resolved (4D) reconstructions that represent organ or patient motion. For that purpose, existing approaches typically acquire a motion surrogate signal along with the CBCT scan such that different motion phases can be reconstructed independently. However, as this strategy requires additional patient preparation and may fail if there is too little data in a certain motion phase, we propose a projection-based motion correction that is able to recover any motion state that has occurred during the scan and thereby overcomes the drawbacks of existing approaches.

Material and Methods

A) Projection-based Motion Correction

In case of a moving patient $f(\boldsymbol{r},t)$, the projections of a CBCT scan are given as:

$$q(\theta, u, v) = \mathsf{X} f(\boldsymbol{r}, t), \tag{1}$$

where θ is the view angle, u and v are the detector coordinates and X denotes the X-ray transform operator. Given a set of projections $\{q\}$, motion correction approaches aim to reconstruct a 3D representation of a certain motion state $f_\tau(\boldsymbol{r})$ at time point $t = \tau$. While existing approaches operate in image domain to do so, we rather propose to predict projection images

$$p_\tau(\theta, u, v) = \mathsf{X} f_\tau(\boldsymbol{r}), \tag{2}$$

which can be reconstructed subsequently to derive the desired volume $f_\tau(\boldsymbol{r}) = \mathsf{X}^{-1} p_\tau(\theta, u, v)$. Here, this is realized as depicted in figure 1, via a neural network which is trained to learn the mapping M:

$$M : q(\theta, u, v) \cdot w(\theta, u, v, c) \rightarrow p_\tau(\theta, u, v), \tag{3}$$

where $w(\theta,u,v,c)$ represents a weighting factor that encodes additional depth information.

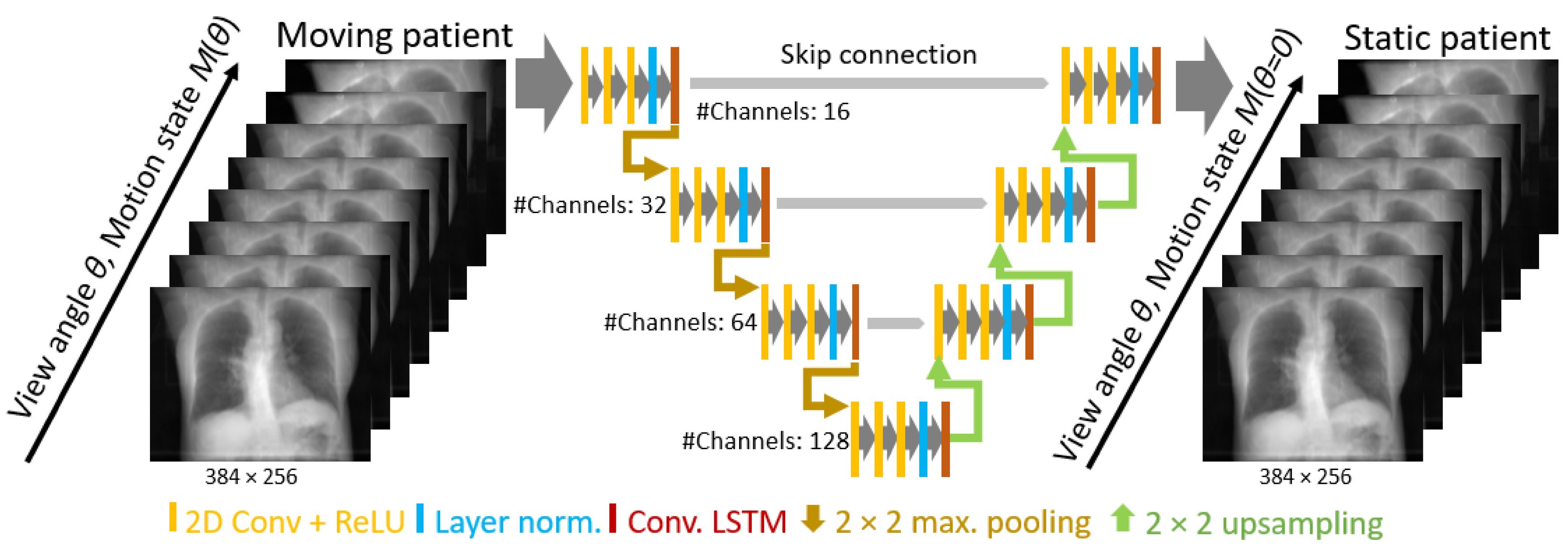

Figure 1: Schematic of the proposed LSTM-based motion correction.

B) Data Generation and Training

Training the proposed approach in a supervised manner requires paired training data, i.e. q's and the corresponding p_τ's. To obtain these data, we rely on CBCT simulations that are based on 55 respiratory-gated 4D CT scans. Each of these scans consists of 10 evenly distributed respiratory phases. Projections q of moving patients are then simulated based on these 4D CT scans according to equation (1) while labels are simulated according to equation (2) by freezing the patient in a certain motion phase τ. Given these data, the network was trained for 300 epochs on an NVIDIA GeForce RTX 3090 by minimizing the mean squared error between the prediction and corresponding ground truth.

Results

The performance of the proposed approach was evaluated for projection images as well as for the corresponding reconstructions of an independent testing data set. Exemplary projections corresponding to a single motion cycle are shown in figure 2. As indicated by the difference images, the proposed approach is able to map all projections to a single motion state.

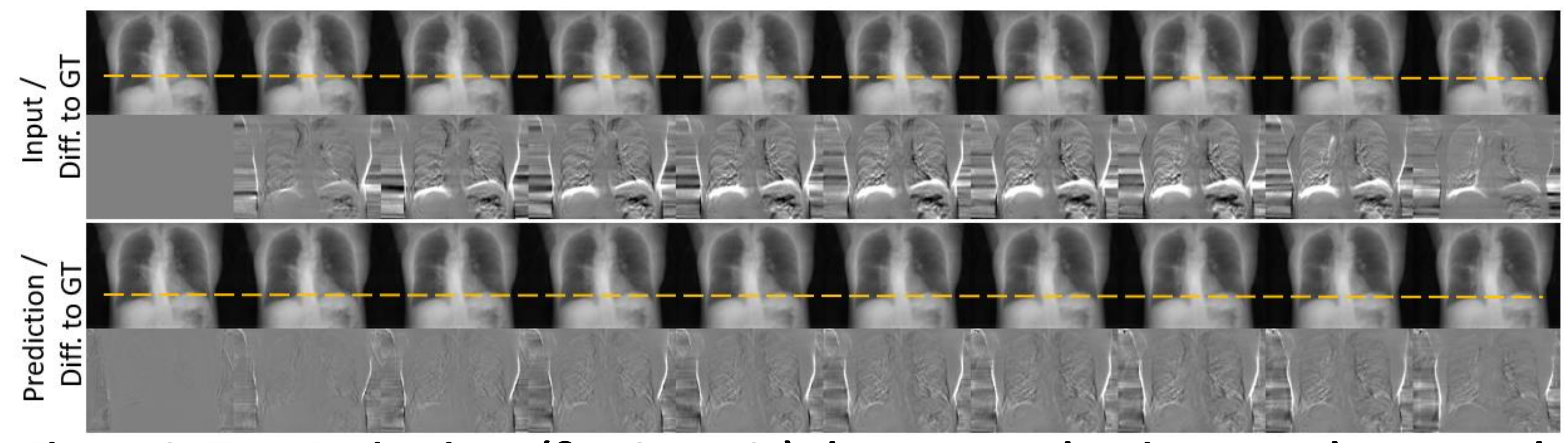

Figure 2: Top: Projections ($\theta = 0^\circ - 10^\circ$) that are used as input to the network and relative difference to the ground truth that keeps the patient frozen in the first motion state (C = 0 %,W = 40 %). Bottom: prediction of the network.

A similar evaluation was performed for reconstructions representing all motion states of the motion cycle. Since the proposed approach is trained to map all projections to the motion state of the first projection, this can be achieved by applying the mapping N times, each time with the n^{th} projection being the first in the sequence. The corresponding results of a test patient in end-exhale and end-inhale are shown in figure 3.

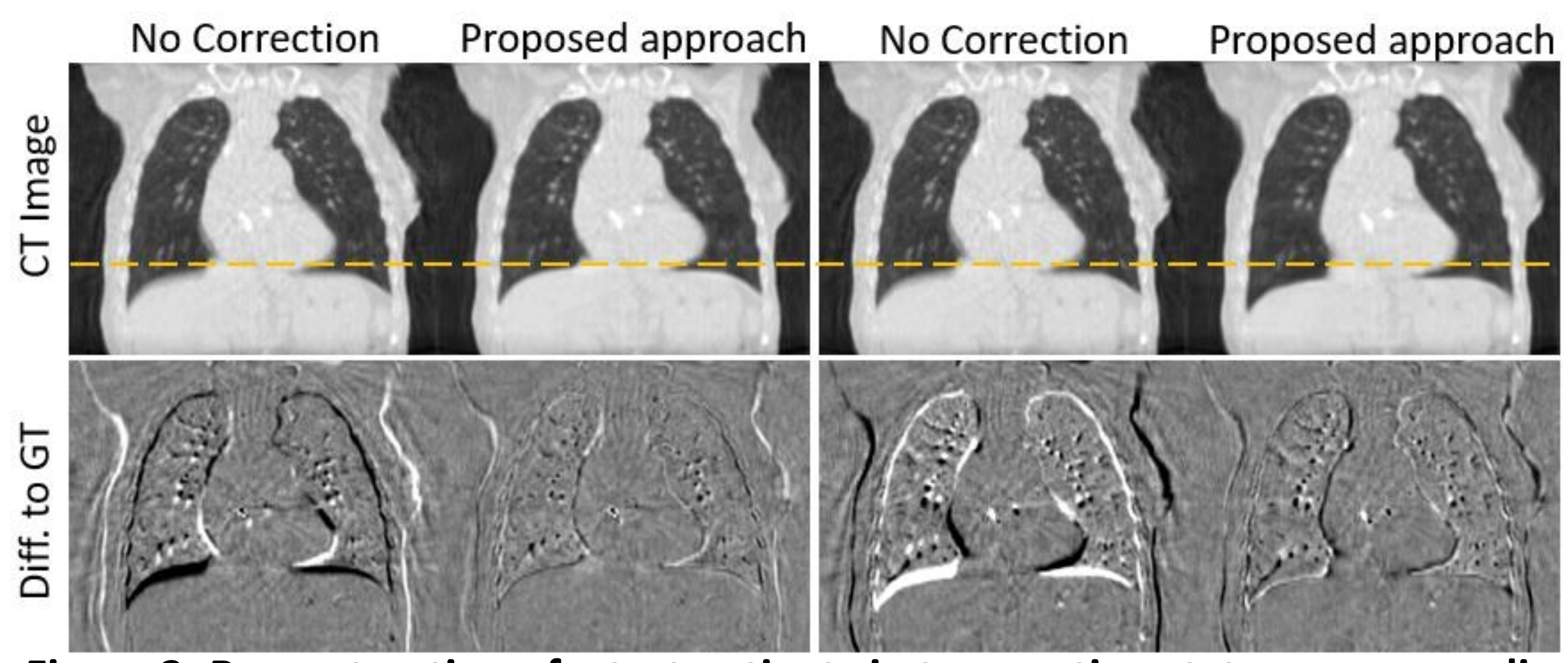

Figure 3: Reconstruction of a test patients in two motion states corresponding to end exhale (left) and end inhale (right).

Discussion and Conclusions

Image-based motion compensation requires the underlying phase-correlated reconstructions to have a certain image quality. However, in case of a highly irregular or sparse angular sampling, this image quality may not be achieved, leading to poor results of the motion compensation algorithm. Here, we circumvent this problem by accounting for motion directly in projection domain. In this way the result is not impaired by reconstruction artifacts and there is no need for additional motion surrogate signals. Our experiments demonstrate a convincing performance in reducing motion artifacts and suggest that convolutional LSTMs are a promising architecture for such an approach.

Correspondence: joscha.maier@dkfz.de