

Model-free cardiorespiratory motion prediction from X-ray angiography sequence with LSTM network

Fariba Azizmohammadi^{1*} Rémi Martin¹ Joaquim Miró² and Luc Duong¹

Abstract—We present a novel model-free approach for cardiorespiratory motion prediction from X-ray angiography time series based on Long Short-Term Memory Recurrent Neural Networks (LSTM). Cardiorespiratory motion prediction is defined as a problem of estimating the future displacement of the coronary vessels in the next image frame in an X-ray angiography sequence. The displacement of the vessels is represented as a sequence of 2D affine transformation matrices allowing 2D X-ray registrations in a sequence. The new displacement parameters from a sequence of transformation matrices are predicted using an LSTM model. LSTM is a particular form of Recurrent Neural Network (RNN) architecture suitable for learning sequential data and predicting time series. The method was developed and validated by simulated data using a realistic cardiorespiratory motion simulator (XCAT). The results show that this method converges quickly and can predict the complex motion in the angiography sequences with irregularities. The mean values of prediction error over all the patients are approximately 0.29 mm (2 pixels) difference for the combination of both motions, 0.51 mm (3.5 pixels) difference for cardiac motion and 0.44 mm (3 pixels) for respiratory motion.

I. INTRODUCTION

Navigation guidance during cardiac interventions, such as balloon angioplasty and stent placement are performed generally under X-ray fluoroscopy [1]. During percutaneous coronary interventions (PCI) catheters and arteries are visualized by X-ray angiography time series. Contrast agent material must be injected to be able to track the arteries during the intervention. It is crucial to minimize the amount of contrast agent injection in order to have a less invasive intervention.

During the cardiac intervention, several organs including the arteries are moving given the heart beating, respiratory movement and sometimes the patient's movements. These movements not only degrade the image acquisition but also make the navigation and guidance more difficult. Hence, there is a demand to compensate the induced motions by estimating and predicting the target's (arteries') movements. Moreover, while the target's movements are tracked in the images, the needs to inject the contrast agent to visualize the vessels will be reduced [2]. Motion prediction can impact robotic-assisted interventions as well, a field in the fast expansion. It is highly required to counteract the systematic latencies through target tracking and mechanical constraints. This compensation can be performed by estimating and predicting the future target positions [3].

*Corresponding author: fariba.azizmohammadi.1@ens.etsmtl.ca

¹École de technologie supérieure ÉTS, 1100 Notre-Dame St W, Montreal, QC H3C 1K3

²Department of Pediatrics, CHU Sainte-Justine, Montreal H3T 1C5, Canada

In recent years, numerous approaches have been developed for controlling the cardiorespiratory motion as well as minimizing its effect. However, challenges still arise regarding a general and not patient-specific model that can compensate not only the cardiorespiratory motion but also the unexpected patient's movements.

Cardiorespiratory motion compensation is relevant for a wide range of interventional procedures involving coronary arteries, pulmonary arteries. Moreover, CT and MR motion compensation protocols are important for proper imaging of anatomical structure under motion. In general, the motion compensation methods are categorized into three groups: model-based, model-free and hybrid approaches. In model-based methods, the motion is represented in a special mathematical mode like linear prediction, Bayesian filtering (Kalman, Extended Kalman, and particle filtering), sinusoidal model, support vector machine and hidden Markov model etc. Motion models use surrogate data as an input and come up with a motion estimation as an output whilst they are patient-specific [4], [5], [6], [7], [8], [9], [10], [11]. The second category includes the model-free methods which are heuristic learning-based algorithms to find a pattern for respiratory motion having a lot of observed data [12], [13], [14], [15], [16]. The current available model-free motion compensation approaches were developed only to predict respiratory motion. Most of these methods outperformed the model-based group. The third category includes methods which are a combination of these two approaches and are called hybrid methods [17], [18].

In this work, we present a novel model-free method using a supervised LSTM network to predict cardiorespiratory motion in angiography sequences. In our approach, the motion signals are extracted from the images and represented by transformation matrices. Then, the LSTM network is trained to predict the next geometrical transformation in the next frame from previous ones.

II. METHODOLOGY

In order to predict the motion behavior, geometrical features are selected and extracted from the 2D images of the X-ray angiography sequences. These geometrical features are represented by transformation parameters (translation, rotation, scaling, and shearing) which are extracted frame by frame in a sequence by an affine registration method. Then, the new position of the arteries in the upcoming frame in the sequence can be predicted using an RNN-based network.

A. Pre-processing

The require steps for extracting the motion features are as follows: At first, some preprocessing and filtering strategies are required to segment and accordingly to extract the center-lines of the arteries. Then the 2D motion feature extraction can be done on the 2D X-ray sequences while registering the X-rays frame by frame in the sequence.

1) *Data simulation with XCAT simulator*: Different motion patterns including different parameters and different patient anatomy were simulated using the 4D XCAT simulator. This simulator generates realistic simulation of cardiac contractions and breathing. One major challenge to address the problem of predicting complex irregular motion signals is the lack of available data. Thus, it is necessary to have an efficient tool which can be used for technique testing, evaluation, and comparison. Each technique involves several selectable parameters for image acquisition, reconstruction, processing, and analysis [22].

In this work, we have used three different patient anatomies with age variety and genders (Table I). For each patient, three different types of motion (only cardiac, only respiratory and both motions) were generated. The simulation was done also in different circumstances in which the patient has normal and abnormal respiratory and heart beating cycles. The normal value for the length of the heartbeat cycle is 1 second and for the respiratory cycle is 5 seconds those can vary among patients and change if the patient is under stress or not breathing normally [23]. The length of the sequences was between 120 to 150 frames to capture at least five heart and/or respiratory cycles.

2) *Segmentation and centerline extraction of the X-rays*: The vascular structure is extracted from the original X-rays and segmented by applying image processing filters such as the Median filter and Frangi filter [19]. Fig. 1 shows the extracted vessel structure from the original X-rays. The Frangi filter parameters are set based on the diameter of the coronary arteries. For sigmas an interval of [1,6] is considered with a step size of 0.1. From the segmentations, the centerlines of the arteries can be extracted using the morphological skeleton operation.

B. Coherent Point Drift (CPD) registration

Point set registration algorithm is widely used in computer vision problems such as image registration. The registration can be rigid or non-rigid. The CPD algorithm is based



Fig. 1. Preprocessing steps on the X-rays to segments the vessels

on Gaussian Mixture Model (GMM) while assigning correspondence points among two sets of points. Then, given the type of registration, it retrieves the transformations for mapping each point set to the other [20]. The two point sets can be aligned and registered rigidly and non-rigidly while considering the alignment as a probability density estimation problem. Then fitting the GMM by centering the first point set to the second and maximizing the likelihood, the GMM is forced to move coherently as a company to retain the topological structure of the point sets [20]. In case of affine registration, a coherence constraint is inflicted by re-parametrizing of the GMM centroid locations with affine transformation parameters (translation, rotation, shearing, scaling). These parameters are concatenated to build the Affine Transformation matrix (AT) as follows:

$$AT = \begin{bmatrix} s_x \cos(\theta) & s_y \sin(\theta) & x - c_x s_x \cos(\theta) - c_y s_y \sin(\theta) \\ -s_x \sin(\theta) & s_y \cos(\theta) & y + c_x s_x \sin(\theta) - c_y s_y \cos(\theta) \\ 0 & 0 & 1 \end{bmatrix} \quad (1)$$

While $A_{00} = s_x \cos(\theta)$, $A_{01} = s_y \sin(\theta)$, $A_{10} = -s_x \sin(\theta)$, $A_{11} = s_y \cos(\theta)$, $T_x = x - c_x s_x \cos(\theta) - c_y s_y \sin(\theta)$ and $T_y = y + c_x s_x \sin(\theta) - c_y s_y \cos(\theta)$. We used notations A_{00} , A_{01} , A_{10} , A_{11} , T_x , T_y for the predicted parameters.

The extracted centerlines of the arteries are considered as bright point sets. We register every centerline point set in each frame to the previous frame in a sequence using CPD algorithm while the registration is affine. However, multiple factors, including an unknown nonrigid spatial transformation (deformation), noise, and outliers caused by the segmentation can reduce the accuracy of point set registration.

C. Recurrent Neural Network (RNN) based time series prediction with LSTM

Long Short-Term Memory (LSTM) is an RNN architecture to solve the vanishing and exploding gradient problem and optimizing the RNNs memory. The LSTM structure compounds memory blocks instead of hidden units in the conventional RNNs. The memory blocks include memory cells which can store the temporal information of the sequential data as well as specific multiplicative units called gates in order to control the flow of information. Each memory block contains an input gate to control the flow of input activations into the memory cell, an output gate to control the output flow of cell activations into the rest of the network and a forgetting gate [21]. Therefore, an LSTM network is able to keep only the necessary information from the past and forget the rest to optimize its memory.

TABLE I
THE VARIATION OF THE SIMULATED DATA.

Patient gender and age	Heart beating cycle	Respiratory cycle
Male-1 year old	1 SEC (Normal)	5 SECS (Normal)
Male-50 years old	1 SEC (Normal)	5 SECS (Normal)
Female-50 years old	1 SEC (Normal)	5 SECS (Normal)
Male-50 years old	2 SECS (Abnormal)	8 SECS (Abnormal)
Male-1 year old	3 SECS (Abnormal)	6 SECS (Abnormal)

To predict a single frame including the affine transformation matrix given the previous frames, we used a many-to-one LSTM structure in which given a sequence of frames as input we are expecting one single frame as output. Fig.2 and Fig.3 show the structure of many-to-one frame prediction.

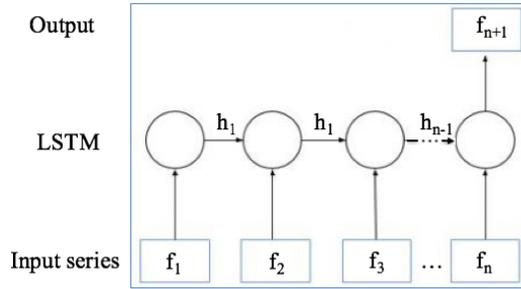


Fig. 2. One-To-Many LSTM structure

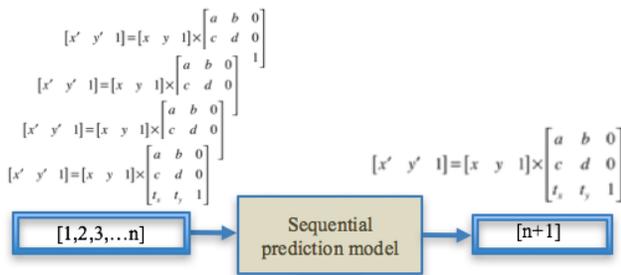


Fig. 3. Affine transformation matrix sequence prediction for the next frame

The LSTM network is trained to predict the arteries transformation in the next frame from the previous ones.

D. New frame prediction in angiography sequences with LSTM

In this section, we explain how LSTM performs the prediction of new transformation parameters in the upcoming frame in a sequence.

1) *Feeding the LSTM network with transformation matrices as inputs and training the LSTM network:* Let $N = 6$ be the number of 2D affine transformation parameters representing translation, rotation, shearing and scaling ($T_x, T_y, A00, A01, A10, A11$), and T is the number of frames or the number of transformation matrices. To effectively feed the LSTM we sort the parameters in a vector X^t of size $N * T$. This vector is called the transformation parameter vector (TP). Then the values in the vector TP are normalized to be fed into the network. The normalization was required since the range of values for some parameters are so small or big and in that case, the network can not learn or converges slowly. Then at the end of prediction, they can be de-normalized to have the real values. Now the prediction problem is defined as solving the predictor of X^t (denoted by \hat{X}^t) via a series of previously measured TP vectors.

We assumed that all the parameters are independent from each other. Thus, to predict the TP vector X^t the model

predicts one element or parameter x_n^t at a time by feeding the LSTM one vector $(x_0^t, x_1^t, \dots, x_n^t)$ of size $n = T$ at a time.

Then given the fact that the real-time prediction of the TP needs continuous feeding inputs and learning, over the time by increasing the number of frames the prediction process becomes slow. Thus, to come up with a solution a learning window W is considered by a fixed number of previous frames to learn from in order to predict the current TP (Fig.4).

The LSTM network is trained using Truncated Backpropagation Through Time (TBPTT) so that the sequence is processed one-time slot at a time and periodically an update is performed back for a certain number of time slots.

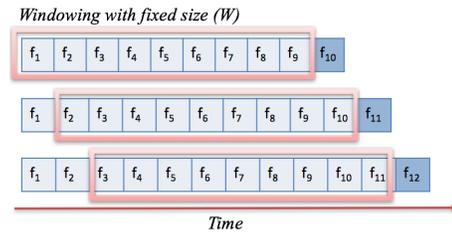


Fig. 4. Sliding learning window

2) *Assessing metric:* We used Mean Absolute Error (MAE) to assess the quality of the predictions because this metric is not overly sensitive to outliers and can simply evaluate the overall error. Given the fact that the segmentation is not perfect, some parts of vessel centerlines in the frames may be lost or got extended by noise. Hence, we had to choose a metric to be less sensitive to this problem while we compare the predicted values to the ground truth resulted in CPD registrations.

Additionally to evaluate the overall error of predicted transformed centerlines we first calculated the distance transform of the original centerlines image. For each pixel of the background, we obtained its distance to the closest centerline point. The distance transform or distance field for each white pixel on the extracted centerline assigns a number that is the distance between that pixel and the nearest nonzero pixel of the vessels. Thus, to calculate the final distance we projected the predicted transformed centerline on the distance transform matrix and averaged the obtained values as an overall prediction error.

III. EXPERIMENTS AND RESULTS

We have used three different patients simulated in normal and abnormal modes while having 120 to 150 frames for 3 different motions (cardiac, respiratory and both motions). The vessel centerlines were segmented and extracted using Frangi filter and skeleton for the all frames of each sequence. Then to extract the motion features we applied affine registration using CPD algorithm. For each sequence, a set of transformation matrices representing the motion features are used as the input for an LSTM network. The transformation matrices include 6 parameters representing translations (T_x, T_y) and rotation, shearing, scaling ($A00, A01, A10, A11$) in 2D.

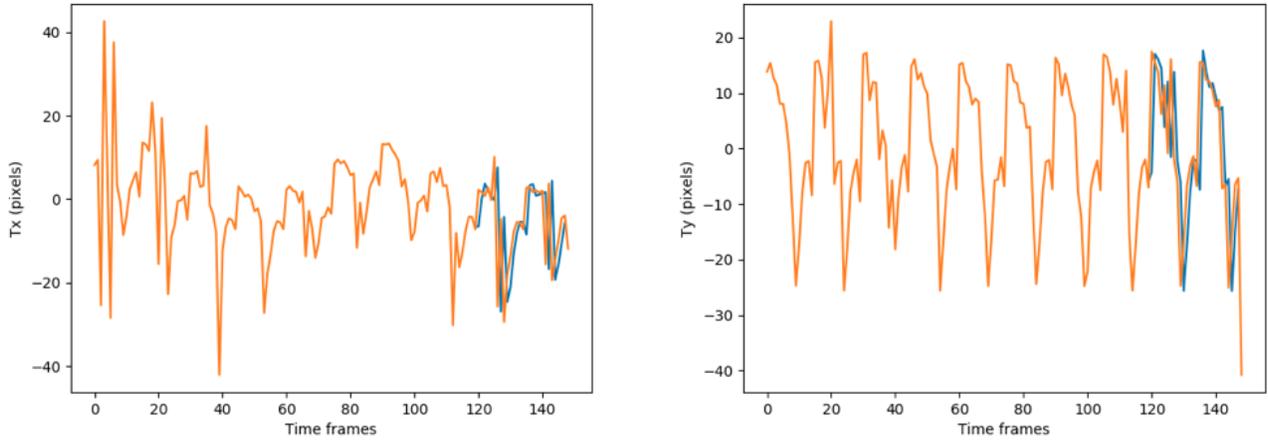


Fig. 5. Prediction of 2D translation parameters for moving arteries with both cardiac and respiratory motions. The ground truth values are shown in orange color while the blue lines show the predictions

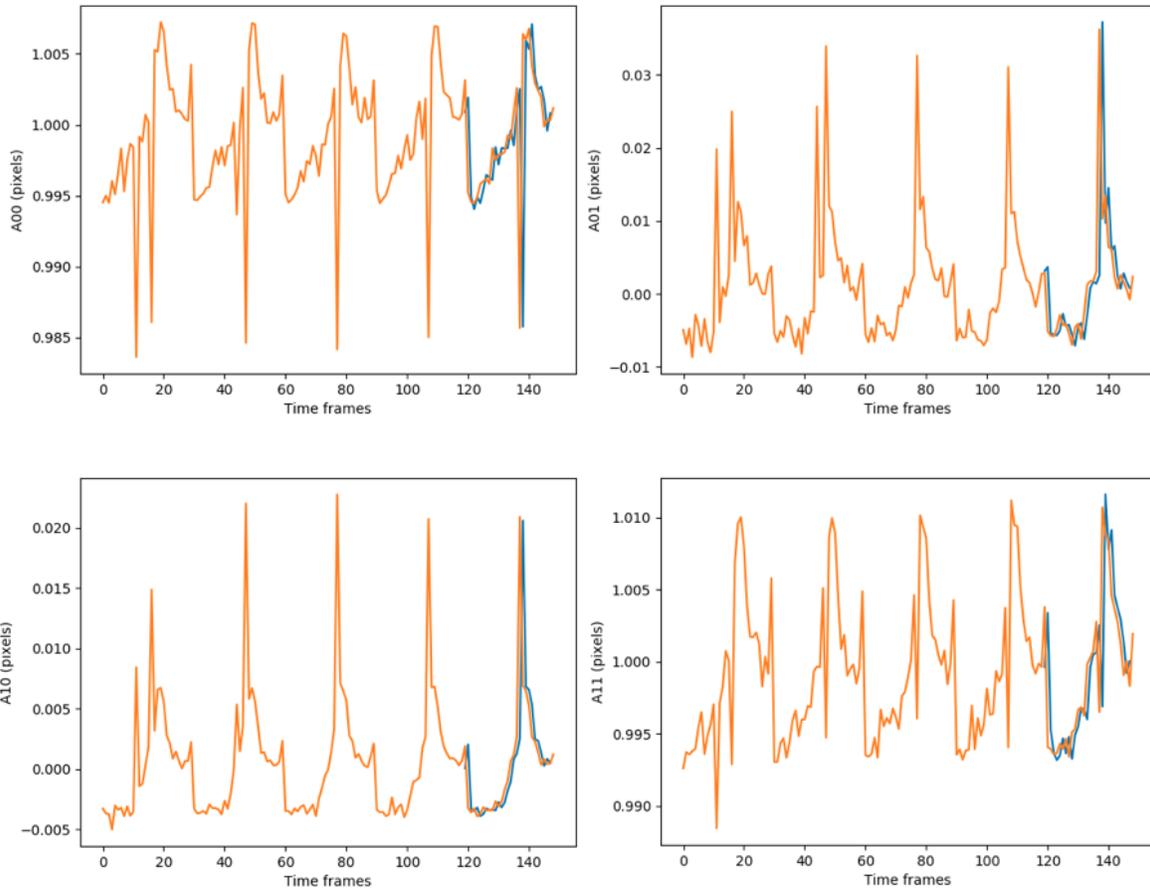


Fig. 6. Prediction of 2D affine transformation parameters (rotation, shearing and scaling) for moving arteries with only cardiac motion. The ground truth values are shown in orange color while the blue lines show the predictions

These values are normalized by dividing by the maximum value in each TP vector. Since the prediction is considered as a regression problem we used a linear activation function for our model and the RMSProp as an optimizer for compiling. Finally, each TP vector was predicted separately while 80 percent of each TP vector was considered as the training set

and 20 percent as the testing set.

Based on the experience, the best number of epochs between a range of (100 to 1000) was 200 epochs and we set that number for predicting the entire values. Keras library was used for building and training the model. The accuracy of the method is evaluated by comparing the predicted values

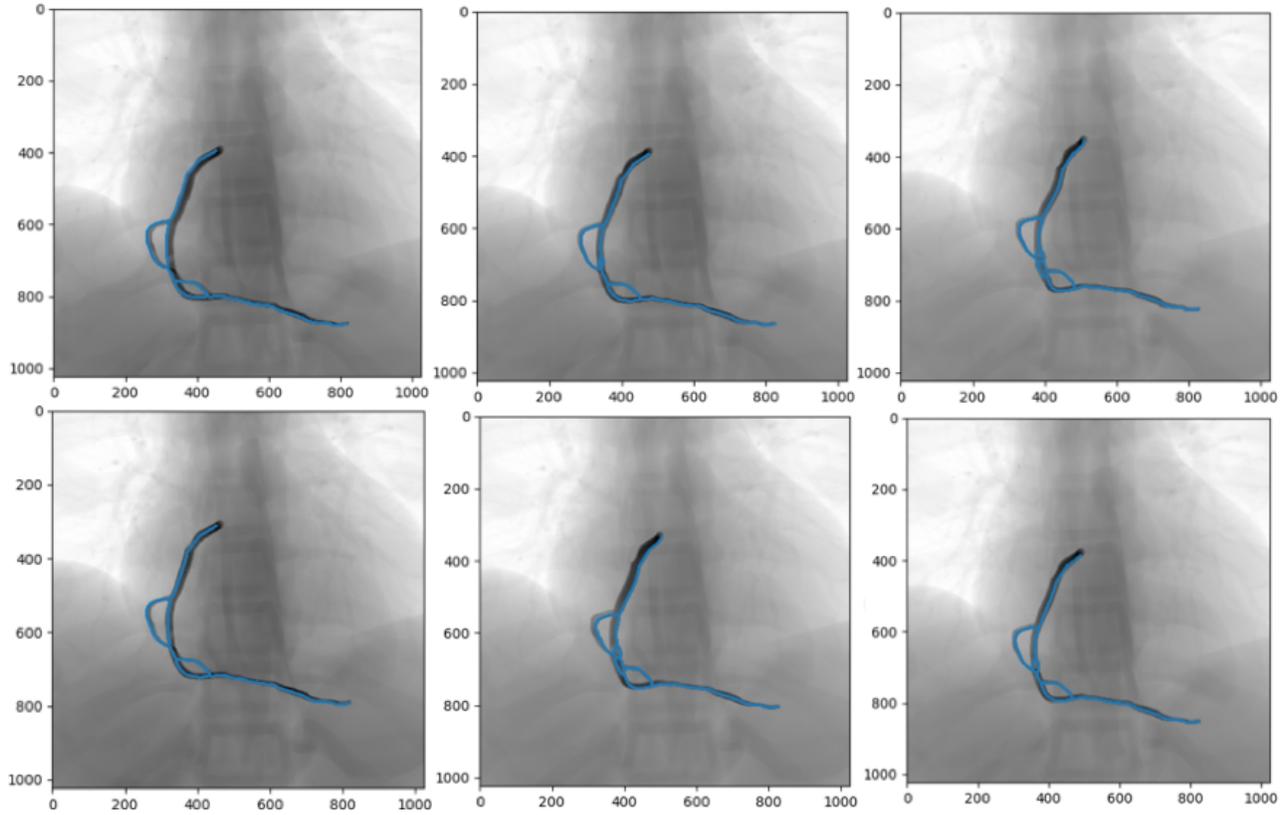


Fig. 7. Overlaying the transformed vessels with predicted transformation parameters (blue colored vessel) with the original transformed images

TABLE II

MAE ERROR FOR PREDICTING THE TRANSFORMATION PARAMETERS FOR ONLY CARDIAC AND BOTH MOTIONS)

MAE	Both Motions						Cardiac Only					
	Tx	Ty	A00	A01	A10	A11	Tx	Ty	A00	A01	A10	A11
Mean	0.09	0.1	0.11	0.06	0.1	0.09	0.17	0.15	0.13	0.17	0.17	0.19
Max	0.38	0.51	0.75	0.41	0.6	0.48	0.86	0.68	0.36	0.8	0.95	0.78
Min	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

TABLE III

MAE ERROR FOR PREDICTING THE TRANSFORMATION PARAMETERS ONLY RESPIRATORY MOTION

MAE	Respiratory only					
	Tx	Ty	A00	A01	A10	A11
Mean	0.07	0.16	0.14	0.2	0.19	0.17
Max	0.32	0.72	0.79	0.73	0.82	0.32
Min	0.00	0.00	0.01	0.00	0.00	0.00

to first the results of the CPD registration using MAE. Then overlaying the transformed centerlines of the vessels with LSTM prediction on the distance transform of the extracted centerlines from original images.

Therefore, we first evaluate how far our predictions were from the expected values to the estimated ground truth (CPD registration values)(Table II and III) and then we compare our results to the original images by applying the CPD registration on the original extracted centerlines and predicted ones (Table IV). While the parameters are predicted separately with a very low amount of MAE error (Tables II and III), the transformation matrix prediction can

be made by concatenating the predicted parameters in a matrix form and apply it to transform the original images. We obtained a low accumulated error for the prediction of the transformation matrix using the distance transform of the original segmented vessels. Fig.7 shows the overlay of the transformed segmented vessels with predicted transformation parameters and the original transformed images.

Fig.5 shows the results for the prediction of parameters T_x and T_y with combination of cardiac and respiratory motions. The input signal for T_x has irregularities and does not look periodic while the predictions with the LSTM network is close to the ground truth. Also Fig.6 shows the prediction of

TABLE IV

THE AVERAGE OVER ALL SAMPLES DISTANCE TRANSFORM ERROR OF THE ORIGINAL CENTERLINE IMAGE TO THE PREDICTED TRANSFORMED ONE IN MM

Mean DT Error(mm)	Both	Cardiac	Respiratory
	0.29 mm	0.51 mm	0.44 mm

other parameters for both motions.

IV. CONCLUSION

In this paper we have shown that a RNN based network can be used to predict the cardiorespiratory motion signal extracted from 2D X-ray images in angiography sequences. The prediction is based on the geometrical features of the motion represented by transformation parameters in the sequences. However, the accuracy of the prediction indirectly depends on the accuracy of the segmentation and registration algorithm in preprocessing steps. The method can generate a good approximation of transformation parameters.

Based on the achieved results, our LSTM model is able to predict the cardiac motion, respiratory motion as well as complex motion signals including both cardiac and respiratory movements even with irregularities in the signal with low amount of error 0.29 to 0.51 mm.

Although, our predictions for the affine transformation parameters were in the same order of magnitude of the simulated transformation values, the deformation of the vessels among the cardiac and respiratory movements was not taken into account and reduced the accuracy of motion tracking. For the future work, we are planning to evaluate our proposed approach in 3D by applying a 3D/2D registration on actual patient data and compare against simulated data.

ACKNOWLEDGMENT

This work was supported by NSERC Discovery grant. The Titan Xp used for this research was donated by the NVIDIA Corporation.

REFERENCES

- [1] Doss, Mohan. "The importance of adaptive response in cancer prevention and therapy." *Medical physics* 40.3 (2013).
- [2] Baka, N, Lelieveldt, B.P.F, Schultz, C, Niessen, W.J, van Walsum, T.W. (2015). Respiratory motion estimation in X-ray angiography for improved guidance during coronary interventions. *Physics in Medicine and Biology*, 60(9),36173637.doi:10.1088/0031-9155/60/9/3617
- [3] Ernst, Floris Dürichen, Robert Schlaefel, Alexander Schweikard, Achim. Evaluating and comparing algorithms for respiratory motion prediction. *Physics in medicine and biology*. 58. 3911-3929. 10.1088/0031-9155/58/11/3911.(2013).
- [4] R. Wernera, J. Ehrhardt, R. Schmidt, H. Handels, Patient-specific finite element modeling of respiratory lung motion using 4D CT image data. *Med. Phys.* 36(5), 15001510 (2009)
- [5] D. Ruan, P. Keall, Online prediction of respiratory motion: multi-dimensional processing with low-dimensional feature learning. *Phys. Med. Biol.* 55(11), 30113025 (2010)
- [6] N. Riaz, P. Shanker, R. Wiersma, O. Gudmundsson, W. Mao, B. Widrow, L. Xing, Predicting respiratory tumor motion with multi-dimensional adaptive filters and support vector regression. *Phys. Med. Biol.* 54(19), 57355748 (2009)
- [7] A. Kalet, G. Sandison, H. Wu, R. Schmitz, A state-based probabilistic model for tumor respiratory motion prediction. *Phys. Med. Biol.* 55(24), 76157631 (2010)
- [8] M. Schneider, H. Sundar, R. Liao, J. Hornegger, and C. Xu. Model-based respiratory motion compensation for image-guided cardiac interventions. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 29482954, June 2010. doi:10.1109/CVPR.2010.5540038.(2010).
- [9] David P Gierga, Johanna Brewer, Gregory C Sharp, Margrit Betke, Christopher G Willett, and George T Y Chen. The correlation between internal and external markers for abdominal tumors: implications for respiratory gating. *International journal of radiation oncology, biology, physics*, 61 5:15518, (2005).
- [10] Bo-Hwan Jung, Byoung-Hee Kim, and Sun-Mog Hong. Respiratory motion prediction with extended kalman filters based on local circular motion model. *International Journal of Bio-Science and Bio-Technology*, 5 (1):5158, (2013).
- [11] Floris Ernst and Achim Schweikard. Forecasting respiratory motion with accurate online support vector regression (svrpred). *International journal of computer assisted radiology and surgery*, 45:43947, (2009).
- [12] M.J. Murphy, D. Pokhrel, Optimization of an adaptive neural network to predict breathing. *Med. Phys.* 36(1), 4047 (2009)
- [13] G.C. Sharp, S.B. Jiang, S. Shimizu, H. Shirato, Prediction of respiratory tumour motion for real-time image-guided radiotherapy. *Phys. Med. Biol.* 49, 425440 (2004)
- [14] S.S. Vedam, P.J. Keall, A. Docef, D.A. Todor, V.R. Kini, R. Mohan, Predicting respiratory motion for four-dimensional radiotherapy. *Med. Phys.* 31(8), 22742283 (2004)
- [15] M.J. Murphy, S. Dieterich, Comparative performance of linear and nonlinear neural networks to predict irregular breathing. *Phys. Med. Biol.* 51(22), 59035914 (2006)
- [16] D. Pokhrel M.J. Murphy. Optimization of an adaptive neural network to predict breathing. *Physics in medicine and biology*, pages 4047, 2009.
- [17] M. Kakar, H. Nyström, L.R. Aarup, T.J. Nttrup, D.R. Olsen, Respiratory motion prediction by using the adaptive neuro fuzzy inference system (ANFIS). *Phys. Med. Biol.* 50(19), 47214728 (2005).
- [18] D. Putra, O.C. L. Haas, J.A. Mills, K. J. Bumham, Prediction of tumour motion using interacting multiple model filter. *International Conference on Advances in Medical, Signal and Information Processing* (2006). pp. 14.
- [19] Alejandro F. Frangi, Wiro J. Niessen, Koen L. Vincken, and Max A. Viergever. Multiscale vessel enhancement filtering, pages 130137. Springer Berlin Heidelberg, Berlin, Heidelberg, (1998).
- [20] M. Andriy, and X. ong. "Point set registration: Coherent point drift." *IEEE transactions on pattern analysis and machine intelligence* 32.12 (2010): 2262-2275.
- [21] Azzouni, Abdelhadi and Guy Pujolle. A Long Short-Term Memory Recurrent Neural Network Framework for Network Traffic Matrix Prediction. *CoRR* abs/1705.05690 (2017): n. pag.
- [22] Segars, W. P., et al. "4D XCAT phantom for multimodality imaging research." *Medical physics* 37.9 (2010): 4902-4915.
- [23] J.Richard Jennings, Maurits W van der Molen, Riek J.M Somsen, Changes in heart beat timing: reactivity, resetting, or perturbation?, *Biological Psychology*, Volume 47, Issue 3, 1998, Pages 227-241, ISSN 0301-0511,