Motion Prediction of Beating Heart Using Spatio-Temporal LSTM

Wanruo Zhang, Guan Yao, Bo Yang, Wenfeng Zheng, Chao Liu

Motion Prediction of Beating Heart Using Spatio-Temporal LSTM

Wanruo Zhang, Guan Yao, Bo Yang, Wenfeng Zheng, and Chao Liu, *Senior Member, IEEE*

***Abstract*—In robot-assisted cardiac surgery, predicting heart motion can help improve the operation accuracy and safety of surgical robots. Different from the conventional prediction schemes which model the point of interest (POI) with only temporal correlation of past observations, this paper proposes an LSTM-based method by exploiting the spatio-temporal correlation of the 3D movements of POI and auxiliary points (APs) on the same surface of the heart. Three different LSTM models are investigated. The first two models define the POI prediction as a pure time-series forecasting problem based on past POI trajectory, and the third model combines the past observations of POI and new observations of APs to take into consideration the extra spatial correlations for prediction. Experimental comparison studies based on 3D coordinates obtained from real stereo-endoscopic videos demonstrate the superior performance of the proposed spatio-temporal LSTM model.**

***Index Terms*—Motion prediction, LSTM, beating heart, robotic surgery, spatio-temporal correlation.**

I. Introduction

In RECENT years, robot technology has been increasingly used to break the bottleneck of manual minimally invasive surgery (MIS). However, as indicated by Mountney *et al.* [1], robot-assisted MIS has been developing slowly in some complex surgical scenarios due to the challenges brought by the dynamic surgical environment. A typical example is the *off-pump* coronary artery bypass graft surgery (CABG). The off-pump surgery avoids damages and side-effects to patients caused by using heart-lung machines (on-pump). However, operation on a beating heart is very challenging because the rapid heart movements are difficult to deal with manually [2], and the teleoperation mode adopted by surgical robot systems further increases the difficulty of the motion compensation.

Manuscript received October 8, 2021; revised February 13, 2022; accepted February 14, 2022. Date of publication February 24, 2022; date of current version March 28, 2022. This work was jointly supported by the Sichuan Science and Technology Program under Grants 2021YFQ0003 and 2021YFS0015 and the Fundamental Research Funds for the Central Universities under Grant ZYGX2019J059. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Arash Mohammadi. *(Corresponding author: Bo Yang.)*

Wanruo Zhang is with the Glasgow College, University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, China, and also with the Key Laboratory of Machine Perception, Shenzhen Graduate School, Peking University, Beijing 100871, China (e-mail: wanruo_zhang@qq.com).

Guan Yao, Bo Yang, and Wenfeng Zheng are with the School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: yaoguan@std.uestc.edu.cn; boyang@uestc.edu.cn; winfirms@uestc.edu.cn).

Chao Liu is with the Department of Robotics, LIRMM, UMR5506, University of Montpellier-CNRS, 34095 Montpellier, France (e-mail: liu@lirmm.fr).

Digital Object Identifier 10.1109/LSP.2022.3154317

Active motion compensation (AMC) technology has been put forward to improve the maneuverability of surgical robots on dynamic soft tissues such as the beating heart [3]. By tracking the point of interest (POI) on the tissue surface, the operating robot can actively compensate for the physiological motion of soft tissue and thus create a virtually stable operating environment for surgeons. Two fundamental issues, motion *measurement* and *control*, are involved in the AMC, and the motion prediction of POI plays an essential role in addressing both issues [2]. For recursive tracking schemes [4]–[8] that are usually adopted for efficient POI measurement, once the tracking chain is interrupted (often due to dynamic effects in complex MIS, e.g., motion blurring or instrument occlusions), it is difficult to recover on its own. In this case, a well-designed motion prediction algorithm can bridge the interrupted tracking and provide proper initialization for the forthcoming uninterrupted sampling period [2]. Motion prediction is also crucial for the control synthesis of AMC systems, which often resort to predictive techniques to handle heart motion of high bandwidth, especially with measurement sensors of significant time delay or slow sampling rates, e.g., ultrasound imaging [9].

Traditional time-series prediction technologies have been used for the POI prediction, such as the Taken's theorem (TT) based methods [10], [11], vector autoregressive (VAR) [12], [13], and Kalman filter-based methods [2], [14]. Most of the existing methods predict the motion of POI from its past trajectory by assuming that the motion is periodic or quasi-periodic and, therefore, can be modeled using the temporal correlation of past observations. However, from a clinical point of view, it is difficult to predict the long-term behavior of dynamic tissues solely based on time-dependent methods because during the prediction phase the *state* of the prediction model cannot be updated with new observations, and the *state deviation* will accumulate as the number of prediction steps increases. As shown in our previous work [2], the temporal-correlated methods, including the TT, VAR, extended Kalman filter (EKF), and dual Kalman filter (DKF), usually achieve acceptable results for short-term predictions or for phantom hearts with regular periodic movements but perform unsatisfactorily for long-term predictions, especially for real beating hearts which exhibit highly dynamic quasi-periodic movements.

Recently recurrent neural network (RNN), a popular deep learning technology, has shown significant advantages in processing time-series data [15]–[17]. To address the long-term dependence problem [18] that is inherent in standard RNN, a variant RNN, so-called *long short-term memory* (LSTM) [19], has been developed and widely used in regression and prediction tasks such as audio-noise power spectral density estimation [20] and pedestrian trajectory prediction [21]. The RNN-based methods represent new promising and efficient solutions for

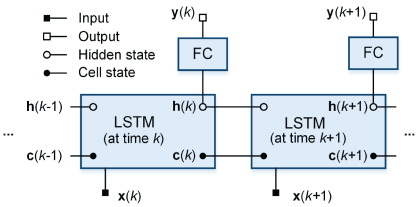


Fig. 1. Unfolded base prediction network.

time-dependent tasks in general. This type of methods fits well in addressing the prediction problem of the beating heart thanks to the powerful ability to model nonlinear dynamic systems. However, to the best of the authors' knowledge, the RNN or its variants have not been explored for the beating heart motion prediction so far.

In this letter, the LSTM-based prediction is investigated and verified based on motion data measured from stereo-endoscopic videos which were recorded through the da Vinci (Intuitive Surgical Inc.) robot. Although the network structure is rather simple, composed of a single-layer LSTM followed by a fully connected (FC) regression layer, the results obtained through this work are encouraging and have important implications for predicting beating heart motion and thus improving the maneuverability of surgical robots on dynamic soft tissues. The main contributions of this work are summarized as follows: 1) The feasibility of modeling heartbeat with simple LSTM networks is verified; 2) A spatio-temporally correlated LSTM prediction model is proposed, which for the first time exploits the spatial correlation between multiple points on the same heart surface to improve heartbeat motion predictions.

II. Methodology

This work aims to predict the future location of a POI on the surface of a beating heart from its past observations (3D coordinates). Three LSTM prediction models, depending on the input and output, are developed: 1) single-point input single-step prediction model, denoted by 1-step LSTM, 2) single-point input multi-step synchronous prediction model, denoted by *M*-step LSTM, and 3) multi-point input spatio-temporal joint prediction model, denoted by ST-LSTM.

The same base network, a single-layer LSTM + an FC regression layer, is adopted by all three models for a fair comparison, which is unfolded in time dimension as shown in Fig. 1. Given an input $\mathbf{x}(k)$ at time $k \in \mathbb{N}$, the state of the LSTM module can be updated as

$$\mathbf{h}(k),\, \mathbf{c}(k) = \text{LSTM}\left(\mathbf{x}(k),\, \mathbf{h}(k-1),\, \mathbf{c}(k-1)\right), \quad (1)$$

where $\mathbf{h}(k)$ and $\mathbf{c}(k)$ are the hidden and cell states, respectively (for more details, please refer to Olah's blog [22]).

The FC layer is a purely linear transformation with a trainable weight matrix, outputting an estimate

$$\mathbf{y}(k) = \text{FC}\left(\mathbf{h}(k)\right). \quad (2)$$

For concision, the base network can be represented in an end-to-end form by hiding its states as

$$\mathbf{y}(k) = \text{Model}\left(\mathbf{x}(k)\right) \quad (3)$$

A. 1-Step LSTM

Given *K* past observations $\{\mathbf{x}(k)\}_{k=1}^{K}$, the 1-step model can be trained by defining the loss function as the sum of squared errors between the estimated and real-observed 3D positions,

$$Loss = \sum_{k=1}^{K-1} \|\mathbf{y}(k) - \mathbf{x}(k+1)\|_2^2, \quad (4)$$

where the output of the model at time *k* is the estimate at the next time, i.e., $\mathbf{y}(k) = \hat{\mathbf{x}}(k+1) \in \mathbb{R}^3$.

Letting $*$ denote a trained model, the single-step predicting at the starting point *K* is written as

$$\mathbf{y}(K) = \text{Model}^*\left(\mathbf{x}(K)\right). \quad (5)$$

For continuous multi-step predicting, the cyclical pattern is adopted:

$$\mathbf{y}(k) = \text{Model}^*\left(\mathbf{y}(k-1)\right) \text{ for } k > K \quad (6)$$

Being trained to match the position at the next step, the 1-step LSTM can usually achieve good short-term prediction performance. For prediction shorter than 1 step, e.g., serving a predictive controller, interpolation techniques can be used to increase the sampling rate of observed signals. On the other hand, it is difficult for 1-step models to handle long-term forecasts because the prediction errors will accumulate rapidly as the number of prediction cycles increases.

B. M-step LSTM

To alleviate the error accumulation problem, *M*-step LSTM is developed by extending the output to the future *M* steps,

$$\mathbf{y}(k) = (\hat{\mathbf{x}}(k+1), \hat{\mathbf{x}}(k+2), \ldots, \hat{\mathbf{x}}(k+M)) \in \mathbb{R}^{3M}. \quad (7)$$

The loss function for training is formulated as

$$Loss = \sum_{k=1}^{K-M} \|\mathbf{y}(k) - \mathbf{d}(k)\|_2^2, \quad (8)$$

where $\mathbf{d}(k)$ is the truth vector concatenated by *M* observations.

The prediction at the starting point *K* is the same as (5). There is no error accumulation within the first *M* steps due to the synchronous prediction. However, error accumulation cannot be totally eliminated in practice as the number of predicted steps required for bridging tracking chains cannot be determined in advance. The cyclical pattern has to be adopted for longer predictions than *M* steps, though the number of cycles compared with the 1-step model is reduced by a factor of *M*. Let $k = K + nM$ with $n \in \mathbb{N}$, then the cyclical pattern can be written as

$$\mathbf{y}(k) = \text{Model}^*\left(\hat{\mathbf{x}}(k)\right), \quad (9)$$

where $\hat{\mathbf{x}}(k)$ is the *M*-th estimate in $\mathbf{y}(k-M)$.

As the limited network resources are balanced for fitting multiple steps, the short-term performance of the *M*-step model (corresponding to the estimation of the first several outputs) may be compromised. Besides, the training sequence available is actually shortened, in which only the first $K - M$ observations can be input for training. Consequently, the *M*-step model usually requires more network resources and training data. Compared with the 1-step model, the gap from the latest training sample $\mathbf{x}(K-M)$ to the prediction start $\mathbf{x}(K)$ is widened to *M* steps.

This means that there is a high probability that the state at the start of the prediction will deviate from the states in training and therefore lead to significant prediction errors even within short steps.

C. ST-LSTM

Both the aforementioned single and multi-step models utilize the temporal correlation within the POI sequence data to estimate the future motion, which can be viewed as the nonlinear extensions of the VAR models [12], [13]. Good performance may be achieved by such models for stationary or less dynamic quasi-periodic movements. However, in a real dynamic setting, it is unrealistic to predict long-term trends based only on the temporal correlation, considering the error accumulation. Tuna *et al.* [13] suggested that physiological signals (e.g., ECG and RPS) may be incorporated for robust prediction. However, this idea faces challenges from multimodal data acquisition and synchronization. Moreover, the correlation between the POI movement and physiological signals remains unclear, and there are some motion components that physiological signals cannot adequately explain. The effectiveness of physiological signals in improving prediction accuracy and coping with dynamic effects such as arrhythmias remains unproven.

We propose a more practical approach by introducing the auxiliary points (APs) on the same surface area of the heart to deal with the error accumulation caused by the long-term blocking of POI observations. There are obvious spatial correlations between different points on the same soft-tissue surface according to soft tissue characteristics. In MIS, the dynamic effects that interrupt the POI measurement are usually concentrated around the POI, so some stable APs without being blocked can always be found whose observations can be used to update the state of the predictive model in time. In addition, the POI and APs can be easily measured synchronously with a visual tracker from stereo-endoscopic images [5]–[7], [23], [24]. Artificial markers can also be fixed on beating heart surfaces, as APs, for robust tracking.

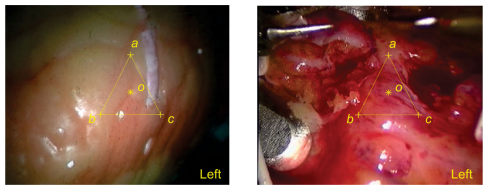
Assuming the number of used APs is N, the input of the ST-LSTM is extended to $\mathbb{R}^{3(N+1)}$ by concatenating the observations of the POI and APs. Considering that the state can be updated in time by the latest observations of APs, the ST-LSTM adopts the single-step prediction scheme, i.e., $\mathbf{y}(k) = \hat{\mathbf{x}}(k+1)$.

In the prediction phase, the cyclical pattern is only applied to the POI, written as

$$\mathbf{y}(k) = \text{Model}^{*}\begin{pmatrix}\mathbf{y}(k-1)\\ \mathbf{x}_{APs}(k)\end{pmatrix}, \text{ for } k > K \tag{10}$$

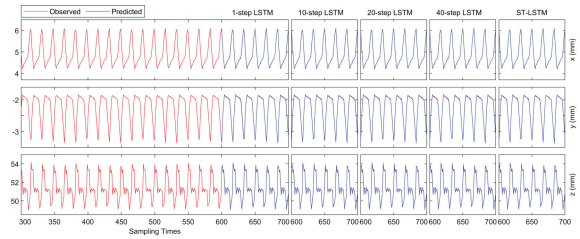
where $\mathbf{x}_{APs}$ is concatenated by the observations of the N APs.

The interpolation techniques and multi-step or skipping prediction schemes can be easily integrated into the ST-LSTM for solving the mismatch of predicted steps caused by too low or too high sampling rates. Since only POI is input cyclically during prediction, the ST-LSTM, with effective training, will learn to rely more on the APs to update its state as the number of predicted steps increases, thus avoiding the error accumulation. In addition, with more inputs, the ST-LSTM has the potential for more accurate short-term prediction than the 1-step model, and this *many-to-one* regression is easier to train.

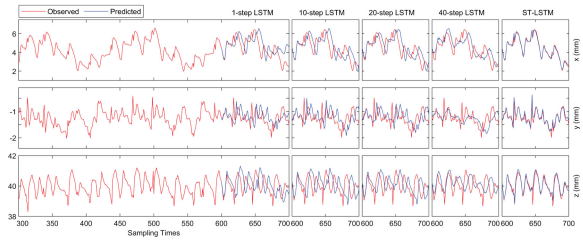


(a) Phantom video. (b) In vivo video.

Fig. 2. The left images of the first frame in stereo-endoscopic videos with the triangle templates for 3D tracking.



(a) Phantom dataset.



(b) In vivo dataset.

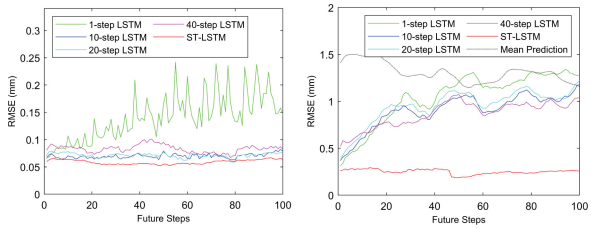
Fig. 3. 100-step prediction results of five models at starting point K = 600.

III. Experimental Results

The 3D motion datasets used for validation were calculated from two stereo-endoscopic videos recorded through the da Vinci robots using a model-based 3D tracking method [7]. Fig. 2 shows the left images of the first frames in these two videos. The phantom video records the movements of a phantom heart model [23], [25], and the in vivo video was captured from a real off-pump CABG surgery [4]. Each video contains 750 stereo frames with a frame rate of 25 Hz. An isosceles triangle region is delineated in the first left image of each video as a template for 3D tracking, where o denotes the POI and is set as the circumcenter of the triangle template, and the vertices $\{a, b, c\}$ are set as the APs. The triangle template is warped with a triangular cubic spline deformation model of 4 control points, corresponding to $\{a, b, c, o\}$, to match pixels at subsequent stereo frames. The 3D coordinates of the 4 control points, corresponding to 12 deformable degrees of freedom, are estimated at each frame with the iterative optimization algorithm in [7]. As a result, a 12×750 motion data matrix is obtained from each video.

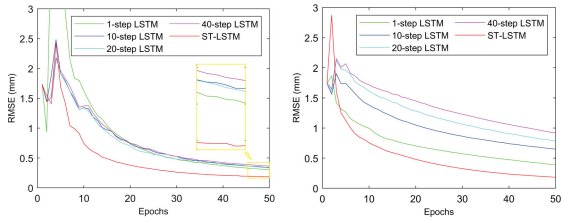
Five models were tested, including the 1-step LSTM, the ST-LSTM, and three M-step LSTM models with $M = 10, 20, 40$. The source codes as well as the motion data for both videos are publicly available on Github[1]. All models employ the same number (300) of hidden state nodes for a fair comparison. Fig. 3 shows the 100-step prediction results of five models at the starting point

[1][Online]. Available: https://github.com/zwr-04/ST-LSTM.



(a) Phantom dataset (b) In vivo dataset

Fig. 4. Prediction errors (RMSE) for five models on two datasets.



(a) Phantom dataset (b) In vivo dataset

Fig. 5. Average learning curves over 1000 trainings on two datasets.

$K = 600$. The predicted curves on the phantom dataset are very close to the observed curves for all five models due to the well-defined periodicity of the phantom heart. In contrast, the motion of the in vivo heart (see Fig. 3(b)) is highly dynamic and thus more difficult to be predicted. The ST-LSTM performs much better than the other four temporal-correlated models, whose predicted curves gradually deviate from the observed curves after about 10 steps, and the cumulative errors are clearly observable.

To avoid assessment biases caused by the specific starting points and the random initialization of model weights, 50 different prediction starting points (from $K = 600$ to 649) were tested for the 100-step prediction, and each starting point was repeated 20 times. The RMSE values over the 50×20 tests at each future step are given in Fig. 4. As a reference, the RMSE values of using the mean of past observations as future estimates were also computed and denoted by *mean prediction*, which indicate the validity of long-term prediction. The validity of short-term prediction can be indicated by the 1-step prediction RMSE of using the last observation as the current estimation, denoted *no prediction*.

On phantom data, all models achieved satisfactory errors (see Fig. 4(a)), which are consistent with the results in Fig. 3(a). All RMSE curves are far lower than the line of *mean prediction* (about 1.6 mm), and the errors at the first step are also lower than the *no prediction* (1.04 mm), which show the effectiveness of the five models for regular motion. The 1-step LSTM yielded fairly good results in the first few steps; however, significant error accumulation occurred in the subsequent steps. The RMSE curves of 10 and 20-step LSTM are very close, and both are better than the 40-step LSTM, which indicates that too large M may lead to poor prediction with the same state resources, and the analyses in Section II-B are hence justified. The results of ST-LSTM are comparatively much better. The total RMSE over $50 \times 20 \times 100$ points is 0.059 mm, even reaching the noise level of visual measurement.

For in vivo data, the 1-step errors of the four temporal-correlated models are all lower than the *no prediction* (0.583 mm). However, their error curves are close to the *mean prediction* line after about 30 steps. It shows that the temporal-correlated models cannot track the real states of the heartbeat. They simply updated themselves recurrently based on the patterns learned from training data, which inevitably led to the accumulation of prediction errors. It can also be seen from Fig. 4(b) that the model with a smaller M value has some advantages in short-term prediction, while with larger M performs slightly better in long-term prediction.

The performance of the ST-LSTM on the in vivo data is superior to those of the other four models, with low error levels and no error accumulation. The total RMSE is 0.252 mm, also significantly lower than the error levels (> 0.6 mm) of the traditional prediction techniques reported in [2], including the TT, VAR, EKF and DKF. It verifies that there are clear and solid spatial correlations between the points on the same soft-tissue surface, and through training, the LSTM can learn how, when and to what extent the spatial correlations from APs should be used to predict the future POI. In terms of single-step prediction, the ST-LSTM also outperforms the 1-step LSTM, thus verifying that the spatial correlation can improve not only long-term but also short-term predictions.

At last, the average learning curves over 1000 tests were compared. Since the numbers of training epochs required by five models are different, only the first 50 epochs are shown in Fig. 5 for trend comparison. All five models used the same optimization algorithm (*Adam*) and hyperparameter settings. The ST-LSTM is easier to train than the other four models by showing the fastest convergence speed on both datasets. The learning curves of the four temporal-correlated models are similar on phantom data but significantly different on in vivo data, which shows that the training on the dynamic in vivo data is more complex and sensitive to the selection of M, where larger M usually leads to harder training.

The average training time is calculated based on our platform (Matlab 2019b at Xeon E3-1231 CPU with 8G Memory and NVIDIA Quadro K620 GPU). On the in vivo data, the ST-LSTM and 1, 10, 20 and 40-step LSTM need 50, 100, 200, 150 and 50 epochs of training, respectively, and the average training time (for 600 frame data) is 4.2, 9.2, 17.1, 13.6 and 4.4 s respectively, which demonstrate the advantages of ST-LSTM in training.

IV. Conclusion

This paper presents a novel spatio-temporal correlated heartbeat prediction method based on LSTM. As far as we know, this is the first work in the literature to predict POI motion by considering the spatial correlation of points on the same heart surface. The LSTM-based prediction models were tested systematically through extensive experiments using phantom and in vivo data recorded through real surgical robot devices. Multi-step prediction presents a fair option for regular phantom data, but the number of synchronized steps and model complexity need to be carefully chosen. For dynamic in vivo data, the single-step LSTM is acceptable for short-term prediction. All temporal-correlated LSTM models performed poorly in long-term prediction, and their prediction errors accumulated rapidly as the number of predicted steps increased. Comparatively, the performance of the spatio-temporal LSTM is highly satisfactory on the phantom heart and encouraging on in vivo heart with stable prediction error curves. As a first study of using RNN techniques for robotic-assisted physiological motion compensation, this work verifies the feasibility of using simple LSTM networks for modelling heart beating and shows superior performance of introducing spatial correlation towards achieving accurate and robust motion prediction in robotic MIS with dynamic operating environment.

REFERENCES

[1] P. Mountney, D. Stoyanov, and G. Z. Yang, "Three-dimensional tissue deformation recovery and tracking," *IEEE Signal Process. Mag.*, vol. 27, no. 4, pp. 14–24, Jul. 2010.

[2] B. Yang, C. Liu, W. Zheng, and S. Liu, "Motion prediction via online instantaneous frequency estimation for vision-based beating heart tracking," *Inf. Fusion*, vol. 35, pp. 58–67, May 2017.

[3] O. Bebek and M. C. Cavusoglu, "Intelligent control algorithms for robotic-assisted beating heart surgery," *IEEE Trans. Robot.*, vol. 23, no. 3, pp. 468–480, Jun. 2007.

[4] D. Stoyanov, G. P. Mylonas, F. Deligianni, A. Darzi, and G. Z. Yang, "Soft-tissue motion tracking and structure estimation for robotic assisted MIS procedures," in *Proc. Med. Image Comput. Comput.-Assist. Intervention*, Palm Springs, 2005, pp. 139–146.

[5] R. Richa, P. Poignet, and C. Liu, "Three-dimensional motion tracking for beating heart surgery using a thin plate spline deformable model," *Int. J. Robot. Res.*, vol. 29, pp. 218–230, Feb. 2010.

[6] B. Yang, W. K. Wong, C. Liu, and P. Poignet, "3D soft-tissue tracking using spatial-color joint probability distribution and thin-plate spline model," *Pattern Recognit.*, vol. 47, pp. 2962–2973, Sep. 2014.

[7] B. Yang, C Liu, K. Huang, and W. Zheng, "A triangular radial cubic spline deformation model for efficient 3D beating heart tracking," *Signal, Image Video Process.*, vol. 11, pp. 1329–1336, Oct. 2017.

[8] B. Yang, C. Liu, W. Zheng, S. Liu, and K. Huang, "Reconstructing a 3D heart surface with stereo-endoscope by learning eigen-shapes," *Biomed. Opt. Exp.*, vol. 9, pp. 6222–6236, Dec. 2018.

[9] M. Bowthorpe and M. Tavakoli, "Generalized predictive control of a surgical robot for beating-heart surgery under delayed and slowly-sampled ultrasound image data," *IEEE Robot. Automat. Lett.*, vol. 1, no. 2, pp. 892–899, Jul. 2016.

[10] T. Ortmaier, M. Groger, D. H. Boehm, V. Falk, and G. Hirzinger, "Motion estimation in beating heart surgery," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 10, pp. 1729–1740, Oct. 2005.

[11] W. K. Wong, B. Yang, C. Liu, and P. Poignet, "A quasi-spherical triangle-based approach for efficient 3-D soft-tissue motion tracking," *IEEE/ASME Trans. Mechatron.*, vol. 18, no. 5, pp. 1472–1484, Oct. 2013.

[12] T. J. Franke, O. Bebek, and M. C. Cavusoglu, "Improved prediction of heart motion using an adaptive filter for robot assisted beating heart surgery," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2007, pp. 509–515.

[13] E. E. Tuna, T. J. Franke, O. Bebek, A. Shiose, K. Fukamachi, and M. C. Cavusoglu, "Heart motion prediction based on adaptive estimation algorithms for robotic-assisted beating heart surgery," *IEEE Trans. Robot.*, vol. 29, no. 1, pp. 261–276, Feb. 2013.

[14] S. G. Yuen, D. T. Kettler, P. M. Novotny, R. D. Plowes, and R. D. Howe, "Robotic motion compensation for beating heart intracardiac surgery," *Int. J. Robot. Res.*, vol. 28, pp. 1355–1372, 2009.

[15] E. Prabhakararao and S. Dandapat, "Attentive RNN-based network to fuse 12-lead ECG and clinical features for improved myocardial infarction diagnosis," *IEEE Signal Process. Lett.*, vol. 27, pp. 2029–2033, 2020.

[16] K. Tan, B. Xu, A. Kumar, E. Nachmani, and Y. Adi, "SAGRNN: Self-attentive gated RNN for binaural speaker separation with interaural cue preservation," *IEEE Signal Process. Lett.*, vol. 28, pp. 26–30, 2021.

[17] B. Gundogdu, B. Yusuf, and M. Saraclar, "Generative RNNs for OOV keyword search," *IEEE Signal Process. Lett.*, vol. 26, no. 1, pp. 124–128, Jan. 2019.

[18] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Trans. Neural Netw.*, vol. 5, no. 2, pp. 157–166, Mar. 1994.

[19] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, pp. 1735–1780, Dec. 1997.

[20] L. Xiaofei, S. Leglaive, L. Girin, and R. Horaud, "Audio-noise power spectral density estimation using long short-term memory," *IEEE Signal Process. Lett.*, vol. 26, no. 6, pp. 918–922, Jun. 2019.

[21] Z. Huang, A. Hasan, K. Shin, R. Li, and K. Driggs-Campbell, "Long-term pedestrian trajectory prediction using mutable intention filter and warp LSTM," *IEEE Robot. Automat. Lett.*, vol. 6, no. 2, pp. 542–549, Apr. 2021.

[22] C. Olah, [Blog] *Understanding LSTM Networks*, 2015. Accessed: Feb. 25, 2021, [Online]. Available: http://colah.github.io/posts/2015-08-Understanding-LSTMs/

[23] D. Stoyanov, M. V. Scarzanella, P. Pratt, and G. Z. Yang, "Real-time stereo reconstruction in robotically assisted minimally invasive surgery," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervention*, Berlin, Germany, 2010, pp. 275–282.

[24] D. Stoyanov, "Stereoscopic scene flow for robotic assisted minimally invasive surgery," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervention*, Nice, France, 2012, pp. 479–486.

[25] P. Pratt, D. Stoyanov, M. Visentini-Scarzanella, and G. Z. Yang, "Dynamic guidance for robotic surgery using image-constrained biomechanical models," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervention*, 2010, pp. 77–85.