Deep Neural Network Approach in Robot Tool Dynamics Identification for Bilateral Teleoperation

Hang Su1, Wen Qi1, Chenguang Yang2, Juan Sandoval3, Giancarlo Ferrigno1 and Elena De Momi1

Abstract—For bilateral teleoperation, the haptic feedback demands the availability of accurate force information transmitted from the remote site. Nevertheless, due to the limitation of the size, the force sensor is usually attached outside of the patient’s abdominal cavity for the surgical operation. Hence, it measures not only the interaction forces on the surgical tip but also the surgical tool dynamics. In this paper, a model-free based deep convolutional neural network (DCNN) structure is proposed for the tool dynamics identification, which features fast computation and noise robustness. After the tool dynamics identification using DCNN, the calibration is performed, and the bilateral teleoperation is demonstrated to verify the proposed method. The comparison results prove that the proposed DCNN model promises prominent performance than other methods. Low computational time (0.0031 seconds) is ensured by the rectified linear unit (ReLU) function, and the DCNN approach provides superior accuracy for predicting the noised dynamics force and enable its feasibility for bilateral teleoperation.

I. INTRODUCTION

Teleoperation is defined as the remote control of a slave manipulator by the human operator [1]. Its application has been adopted in numerous areas. Moreover, a substantial novel interest is driven by medical robot applications, Robot-assisted Minimally Invasive Surgery (RA-MIS), for instance. In particular, bilateral teleoperation [2], [3] draws lots of research interests because it provides tactile feedback for robotic surgery, which can promote the manipulability for the tasks performing and ease of surgical operation, for example, enhancing surgical accuracy [4], optimizing robot dexterity and reducing the injuries of the patient [5], [6]. In decades, it has been proved that it is essential to provide haptic feedback of surgery interaction forces to the surgeons [7]–[9]. Without haptic feedback in minimally invasive surgery, the possibilities of some disadvantageous effects such as intra-operative injury and time-consuming will increase [10]. Consequently, precise force sensing of the surgical tool is crucial in the bilateral teleoperation for Robot-assisted surgery [11]–[13].

Achieving reliable interaction forces in surgery [14] has drawn lots of research interests owing to the limitations imposed by the usage of teleoperation [15], [16]. For applying the interaction forces on the haptic manipulator, a six-axis force sensor is mounted between the slave robot manipulator and the end effector surgical tool [17]. In this situation, the influences of the sensor size and the increment of the equipment cost can be ignored contrary to the case of mounting the force sensor on the surgical tip [18]. However, unexpected disturbances [19], [20] will be introduced due to the dynamics of the surgical tool. Hence, it is significant to determine the tool dynamics force and compensate it online. After this, the calibration of the force direction should be performed to align the force between the sensor and manipulator in the same coordination frame [21].

Previously, several approaches of dynamics identification using robot joint angles have been proposed, based on the kinematic and dynamic models of the manipulator [22]–[25]. A Proportional–Derivative (PD) controller with online gravity compensation utilizing robot joint angles is designed for regular tasks of the robot in [26]. A nonlinear optimization scheme is proposed to obtain feasible values from the identified dynamic coefficients in [27]. However, the methods depending on the kinematic model are challenging for industrial and biomedical robots, for which the joint torque sensors are often unavailable. It forces the researches towards novel approaches for tool dynamics identification. In our previous works, it has been proved that the force of tool dynamics mapped on the force sensor relies on the orientation of the robot [28], [29]. However, it is difficult to obtain an accurate mathematical model for nonlinear regression [30]. Consequently, we utilized curve fitting (CF) and Artificial Neural Networks (ANN) methods, i.e., feed-forward neural network (FFNN) and cascade-forward neural network (CFNN) to identify the tool dynamics based on the orientation angles of the manipulator [28], [29]. Although the ANN-based approaches have generalization ability to build a model without priority information, they have extremely poor robustness which is easily affected by the problems of overfitting and underfitting. The established ANN-based models had no anti-interference ability to predict the tool dynamics forces in the noisy environment.

Recently, various deep learning (DL) methods have emerged to learn the hierarchical feature representations [31], which show prominent performances compared with other general regression methods, such as CF [32], fuzzy logic (FL) [33] and ANN [21]. The benefits including fast computation, noise robustness, and stability, are brought by the DL algorithms, e.g., the recurrent neural network (RNN) and deep convolutional neural network (DCNN). Furthermore,
Fig. 1: Teleoperation scenario in robot-assisted surgery. The surgeon is placed at the remote site to move the haptic manipulator following a desired trajectory $X_r \in \mathbb{R}^m$. In the slave site, the robot manipulator executes the trajectory $X_d \in \mathbb{R}^m$. The transformation matrix, $\psi_m T$, is from the slave to remote coordinate frames. $\psi T$ is the transformation matrix between the slave base and the surgical tip. The matrix, $\psi T$, represents the transformation from the operational space to the camera. A camera is placed in the slave site to enable the visualization of the surgical operation for the surgeon.

compared to the rest of the activation functions, the Rectified Linear Unit (ReLU) method is safe concerning the gradient vanishing problem, less computationally expensive [34] and the dropout technique was used generally to reduce the complexity and to avoid overfitting [35].

In this paper, a novel DCNN-based nonlinear regression model is proposed for the tool dynamics identification to achieve accurate force sensing. It consists of several convolutional networks, ReLU function, and dropout layer for achieving noise robustness, fast computation, and stability enhancement. The proposed DCNN-based regression model is implemented on the real robot application to eliminate the tool dynamics. Finally, a bilateral teleoperation demonstration is performed to verify the proposed DCNN-based tool dynamics identification approach. The novel contributions of this work include:

1) A model-free approach is presented to model the tool dynamics using the tool pose information.
2) A DCNN-based regression algorithm is utilized to achieve tool dynamics identification with fast computation and noise robustness.
3) A bilateral teleoperation demonstration is implemented to verify the effectiveness of the DCNN-based tool dynamics identification method.

II. RELATED WORK

In our previous work, several nonlinear regression approaches are adopted for modeling the tool dynamics using the tool orientation (e.g., Euler angle). The CF and FFNN with single hidden layer methods were implemented for tool identification and the performance were compared for bilateral teleoperation in [28]. As a model-free approach, the modeling ability of FFNN model was more potent than the CF method, which depends on the mathematical model accuracy. Although the used FFNN method with 30 neurons in the hidden layer was examined as a promising method, its stability regularly suffered from either overfitting or underfitting. To assess the best ANN-based model, we compared the accuracy and computational time among several FFNN and CFNN models, which features with different neurons and layers [29]. The best structure is the CFNN model with 9 and 6 neurons in each layer. However, our previous works aimed to map three dimensions Euler angles which limited the accuracy because the tool dynamics includes the position movement. Primarily, it becomes very challenging to process dataset with the movement of the tool. Meanwhile, the limitation also exists when the ANN-based models are applied to the noise dataset.

III. PROBLEM FORMULATION AND METHODOLOGY

In the teleoperated surgery, the master haptic and slave robot have different kinematics characteristics and workspace [36], which requires an efficient and convenient approach to provide accurate workspace mapping to enable the surgeon to access the whole workspace in the patients’ cavity [37], as shown in Fig. 1. Furthermore, the tool dynamics should be identified and compensated to access the accurate interaction force on the surgical tip during the motion. For achieving the tool dynamics identification, DCNN-based modeling techniques have been implemented on the same online procedure to compare its accuracy. Once the tool dynamics identification is achieved, the calibration between the force sensor and robot pose is implemented to map the force into manipulator coordinates, making precise force feedback measurement for the remote site. Finally, a demonstration experiment of bilateral teleoperation is presented to prove the proposed method.

A. Problem Formulation

In order to provide force feedback, helping the surgeon to guide his gestures correctly [38], the sensor is embedded between the robot end-effector and the surgical instrument as presented in Fig. 1. The sensor measures the interaction force
on the tool tip and the tool dynamics at the same time. It is well known that the tool dynamics force can be expressed as [14]:

\[
F = M_{XT}(X_T)\dot{X}_T + h_{XT}(X_T, \dot{X}_T) + G_{tool}
\]

(1)

where \(F\) is the tool dynamics forces, \(X_T\) is the static pose of the tool, \(G_{tool}\) is the tool of gravity force. The matrix \(M_{XT}\) is the inertia matrix of the tool and the matrix \(h_{XT}\) represents the generalized centrifugal and Coriolis influences of the tool.

In our previous works, we only considered the static gravity force without consideration of the full tool dynamics expression, which limits the accuracy of the modeling when there is a movement of the tool. Hence, the inputs of the tool dynamics model in this paper are enriched to 15 dimensions (15D) which includes collected positions \(P = [x, y, z]\), Euler angles \(\theta = [\theta_x, \theta_y, \theta_z]\) and their corresponding velocity, \(V_P, V_\theta\), and acceleration, \(A_P, A_\theta\), defined as:

\[
X = [\theta, V_P, V_\theta, A_P, A_\theta]
\]

(2)

It should be noted that the static position \(P = [x, y, z]\) of the end-effector does not influence the tool dynamics forces. The output is set as the tri-axis force vector \(\hat{F} = [F_x, F_y, F_z]\). The established nonlinear regression model aims to predict the forces on the inputs \(X \in \mathbb{R}^{15}\) with high accuracy as follows.

\[
F = f(X, \Theta)
\]

(3)

The optimal parameters represented by \(\Theta\) can be determined by adjusting the minimum least squares:

\[
\Theta = \arg \min_{\Theta} \frac{1}{N} \sum_{i=1}^{N} (\hat{F}_i - F_i)^2 = \arg \min_{\Theta} \| \hat{F} - F \|_2^2
\]

(4)

where \(\hat{F}_i\) and \(F_i\) are the observed and predicted values, respectively. The Root Mean Square Error (RMSE) shown in Eq. 5 is chosen to estimate the predictive performance of the model.

\[
\varepsilon = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\frac{\hat{F}_i - F_i}{\hat{F}_i})^2}
\]

(5)

To strengthen the assessment, we define the overall RMSE (Eq. 6) to examine the complete regression accuracy of the tri-axis force. Both \(\varepsilon\) and \(\varepsilon^*\) are expected to be 0.

\[
\varepsilon^* = \frac{\varepsilon_{F_x} + \varepsilon_{F_y} + \varepsilon_{F_z}}{3}
\]

(6)

B. DCNN-based Tool Dynamics Identification

1) Training Data Preparation: The 6D raw data \(\{P; \theta\}^T\) (73774 samples) were collected from the workspace (41729 samples for training and 32045 samples for testing) and extended into 15D by Eq. 2. Knowing that the convolutional network is proved to be more valid on the Homogeneous matrix [39], a new input map \(X^* \in \mathbb{R}^{15 \times 5}\) is constructed as:

\[
X^* = [X, X - \bar{X}, (X - \bar{X})/\sigma(X)]^T\}
\]

(7)

\(\bar{X}\) and \(\sigma(X)\) denote the average and variance of \(X\), respectively.

2) DCNN Architecture: Fig. 2 shows the proposed DCNN architecture for mapping the inputs to the 3D forces. Both batch normalization (BN) layer and Rectified Linear Unit (ReLU) activation function are adopted in each convolution module. The former one is to update the network independently while the latter aims to solve the vanishing gradient problem for fast computation. The dropout layer is used for resolving the overfitting. The DCNN-based architecture consists of three main parts as follows.

- Inputs: the size of reconstructed matrix is 15 \times 5
- Deep Convolution Modules: it include four similar deep CNN modules. Each of them has a 2D convolutional layer, a BN layer, and a ReLU function. We adopted four types of window size to perform the convolutional operations, i.e., 8, 16, 32, and 32. All of them used the 2 \times 2 filter. Hence, the size of yielded feature map are 14 \times 4, 13 \times 3, 12 \times 2, and 11 \times 1
- Dropout: to avoid the large convolved feature vectors causing the phenomenon of overfitting and time-consuming, a dropout layer is applied. We set 0.5 as the dropout percentage
- Output: both full connection and regression layers were arranged at the end of DCNN-based structure. The
predicted forces were computed by summing the inputs of the node.

The full connection layer have $3 \times 2400$ weights and 3 bias. The initial learning rate is 0.001. We adopt mini-batches of size 50 with 0.001 drop factor and 300 drop period. The replacement optimization algorithm is adaptive moment estimation (Adam) optimizer. The Mean Squared Error (MSE) is used as the loss function.

3) Noise Robustness: The produced data streams in real-world will be affected by various noises. Hence, to develop a robust model reducing the effect of noise for practical operation is necessary. We added Gaussian noise to the inputs at each epoch for anti-interference measurement. The average and standard deviation of the errors, i.e., RMSEs (Eq. 5) and overall RMSE (Eq. 6), are the two crucial test indicators to search for the best noise robustness model.

C. Force Sensor Calibration

Once the identification of tool dynamics is achieved, the force sensor is calibrated using SVD method [28], [29] to enable force feedback for the teleoperation. Fig. 3 exhibits the overview of the DCNN-based dynamics identification and force calibration procedure.

IV. EXPERIMENTS AND DEMONSTRATION

Experiments are performed to validate the proposed DCNN-based tool dynamics identification approach.

A. System Overview

The description of the developed robot system is shown in Fig. 4. A robot is placed on the remote site to serve as the slave manipulator. Fast Research Interface (FRI) is used to create a real-time access to the robot controller (KRC) at rates of 500 Hz [17]. The controller has been developed using OROCOS 1, with a real-time Xenomai-patched Linux kernel, and using ROS 2 Kinetic under Ubuntu. To guarantee the control frequency, the force sensor, the ROS node and the OROCOS torque controller were executed on separate computers with UDP communication between each other. The developed teleoperation system is integrated with:

- A 7 DoFs slave manipulator (LWR4+, KUKA, Germany) to operate in the slave site.
- A 6-axis force sensor (M8128C6, SRI, China) [40] to measure the interaction force.
- A cable-driven master manipulator (Sigma 7, Force Dimension, Switzerland) for the teleoperation control.

B. Tool Dynamics Identification using DCNN

Based on our previous contributions [28], [29], the anti-noise performance was quantitatively and qualitatively evaluated among CF, ANN-based, and DCNN-based models, i.e., Long Short-Term Memory (LSTM) and DCNN. For ANN-based methods, both two layers and single layers structures are adopted, which are marked as FFNN-M, CFNN-M, FFNN-S, and CFNN-S. The models of multiple layers have 9 and 6 neurons in each layer while the single layers have 20 nodes. The LSTM structure was created with 150 neurons LSTM layer, a dropout layer (0.5 probability), and a full connection layer. It also adopted the Adam optimizer algorithm. The other parameters are set as follows: 200 mini-batch size, 0.2 initial learning rate, 0.5 drop factor, and 100 drop period. We use MATLAB 2018b software (hardware platform: Intel(R) i7 Core, 2.80 GHz CPU, 16.0 GB RAM) to verify the performance of DCNN model.

1) Noise Robustness: The regularization method and the characteristics of parameters decide the noise robustness. Even though the optimal parameters can be obtained in systematic ways, it is not possible for a fair comparison. In this experiment, the Gaussian noise with four types of Signal-to-Noise Ratio (SNR) is annexed into the inputs to

---

1 Open Robotic Control Software, http://www.orocos.org/

measure the anti-interference performance. Seven models are trained based on the above mentioned approaches and the unidentified system (NS) on the training dataset. The comparative errors are calculated on the testing dataset. In each SNR case, it runs 20 times to represent the standard deviation of the computed RMSEs on the four models (displayed in Fig. 5).

In general, both DL-based (DCNN and LSTM) models get better results for predicting $F_y$ and $F_z$ than when adopting ANN-based models. Notably, the DCNN model gets the lowest average error for predicting $F_x$ force. By observing the RMSEs computed by the four methods, the proposed DCNN structure is the most stable method for predicting the noisy data, because it gets the smallest standard deviation compared to the other approaches.

2) Computational Time: The computational time is an important feature for robot tool dynamics identification since force sensors generally require high frame rate to consider the position and orientation changes. The computational time is compared to the testing dataset. We computed the sum
and the average of computational time of ten results among ANN and DNN based models. Meanwhile, both incremental learning (predictive length is 1) and batch learning (predictive length is 5 and 10) methods are used. The computational times (table I) prove that the proposed DCNN model is the faster-speed computation method. In the incremental learning way, it only needs 100.75 seconds to predict all the results, while the LSTM model needs 357.44 seconds. In the two batch learning cases, the DCNN method only needs (less than) 0.0006 seconds to predict the batch results. Although CF model obtains the lowest time, it cannot acquire a satisfied accuracy.

C. Demonstration after tool dynamics identification and calibration

A bilateral teleoperation tracking demonstration with physical interaction is presented. As shown in Fig. 6, the position tracking is achieved when there is a physical interaction on the surgical tip around 70s-75s. When a physical contact occurs, the haptic feedback will prevent the surgeon from continuing moving the haptic manipulator. Hence, the position error is constrained in a small range.

V. CONCLUSION

A model-free approach based on DCNN regression algorithm was proposed in this paper to achieve tool dynamics identification for bilateral teleoperation, which features fast computation and noise robustness. This could meet the clinical requirements in terms of efficiency and accuracy. A calibration procedure was performed after the identification of tool dynamics, and the bilateral teleoperation was demonstrated to verify the effectiveness of the proposed method. In this study, the presented DCNN model improved noise robustness and computation time compared to the conventional approaches, such as CF, FFNN, CFNN, and LSTM. The obtained RMSEs on the noise datasets proved its ability for eliminating perturbations. The DCNN approach provided superior accuracy for predicting the noised dynamics force and enabled its feasibility for bilateral teleoperation.

For future works, more challenging problems will be considered in the bilateral teleoperation control system, such as the dead-zones and time-delays. The system tracking accuracy and stability might not be secured under these conditions, which remain the main preconditions for the success of the surgical procedure. Furthermore, an external force sensor will be placed on the table to validate the real-time contact force acting on the surgical tip.
REFERENCES


