Deep learning for haptic feedback of flexible endoscopic robot without prior knowledge on sheath configuration

Xiaoguo Li, Anthony Meng Huat Tiong, Lin Cao*, Wenjie Lai, Phuoc Thien Phan, Soo Jay Phee

Robotics Research Centre, School of Mechanical & Aerospace Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798

A R T I C L E   I N F O

Keywords:
Flexible endoscopic surgical robots
Tendon-sheath mechanisms
Haptic Force feedback
Force Hysteresis
Deep Learning

A B S T R A C T

Distal-end force information is usually missing in flexible endoscopic robots due to the difficulties of mounting miniature force sensors on their end-effectors. This hurdle creates big challenges in providing a sense of touch for the operating surgeons. Many existing studies have developed models to calculate the distal-end forces based on the measured proximal-end forces of Tendon-Sheath Mechanisms (TSMs), but these models assume known sheath bending configuration which is unknown during real-life surgeries. This paper presents a two-stage data-driven method that makes dynamic distal-end force prediction of a flexible endoscopic robot without this assumption. In stage one, a convolutional neural network is used to estimate the shearh cumulative bending angle based on the proximal-end force responses of the robot to a probing signal; in stage two, a combination of two long-short-term memory models pre-trained for the bending angles nearest to the estimated angle (obtained in stage one) makes dynamic estimations of the distal-end force of the robot. The proposed approach overcomes the challenges due to unknown TSM configurations and can robustly identify the correct force hysteresis phases of TSMs. The force prediction is continuous, accurate, and has a mean RMSE of 0.1711 N. This method was validated on an actual flexible surgical robot. In addition, since the proposed approach provides an estimation of the current system cumulative bending angle, it can also be used to facilitate the mathematical modeling methods which require information on the cumulative bending angle.

1. Introduction

Natural Orifice Transluminal Endoscopic Surgery (NOTES) has received growing attention in the development of Minimally Invasive Surgery (MIS) in recent years [1,2]. Robotic-assisted NOTES offers prominent advantages in tool dexterity, shortened intraoperative time and reduced surgical trauma being done to the patients. A flexible endoscope is used to deliver the robotic arms to the surgical site endoscopically to perform procedures such as dissection and suturing, while the actual housing resides outside the patients’ body [3,4]. In such cases, Tendon-Sheath Mechanisms (TSMs) are often used to transmit the power and movement from the actual housing to the robotic end-effectors through the long and narrow path of an endoscope [5]. However, TSMs are intrinsically associated with nonlinear friction and backlash that pose challenges for accurate position control and haptic feedback. On the other hand, substantial studies have reported that haptic feedback is crucial for the surgeons during robotic-assisted surgeries in aspects such as surgical precision, avoidance of excessive tissue damage and operation fatigue [6–8]. Without the presence of haptic feedback, currently, surgeons have to rely heavily on visual guidance and experience. To acquire the distal-end information for haptic feedback, many miniaturized force sensors have been proposed based on working principles such as resistance [9], displacement [10], current [11], pressure [12], capacitance [13], and optical properties [14]. Among these approaches, optical Fiber Bragg Grating (FBG) force sensors have demonstrated remarkable potential in both the ability to be integrated with TSM as well as its force measurement accuracy. However, the sensor robustness remains the main concern when FBGs are applied in a flexible surgical robot. Extra care needs to be taken to prevent the potential wearing or breakage of the fiber during application. Moreover, none of these electrical force sensors is used in the actual tendon-driven NOTES robot mainly due to the limitations such as sterilization issues, tiny available mounting space at the distal-end, problems with associated wires and accessories, and unstable readings in the harsh working environment during the surgery.

To acquire the end-effector force information without additional sensors, a wide variety of mathematical friction models can be found in the literature that strive to capture the nonlinear relationship between the proximal-end force and the distal-end force. Kaneko et al. developed a lumped mass model using Coulomb friction to estimate the tension loss along the tendon inside the sheath [15]. Building upon analysis presented by Kaneko, Palli and Melchiorri proposed several friction

* Corresponding author.
E-mail addresses: lin.cao@ntu.edu.sg, lin.cao@usask.ca (L. Cao).
models to better characterize the nonlinear hysteresis in TSMs based on Dahl's dynamical friction model [16–18]. However, the complexity of the model increases significantly as more elementary models are being considered. A further effort was made by Agrawal et al. to dynamically predict the backlash, cable slacking and nonlinear friction loss by analyzing the system as many discrete segments [19]. Do et al. [20] proposed dynamic models based on nonlinear hysteresis friction by considering the effect of acceleration and velocity. Later, improved dynamic models [21–23] were proposed by analyzing the hysteresis phenomenon in the sliding and presliding regime, separately, based on Striebeck function and the asymmetric normalized Bouc-Wen model. These models demonstrated an enhanced performance especially when TSM is operating in the vicinity of zero system velocity. However, these mathematical modeling approaches require complex parameter identification processes and ad-hoc optimization strategies such as Nelder-Mead Simplex method and Genetic Algorithm, and they commonly suffer from discontinuity issues. In real-world applications, the above methods are not incorporated in flexible NOTES surgical robots mainly due to two reasons. Firstly, the models are established based on the assumption of constant pretension in the tendon and no system slacking. This assumption does not always hold during the actual operation, therefore the accuracy of the models are not guaranteed. Secondly, the mathematical models assume fixed and known sheath bending angle which is applicable to experimental setups yet it could be difficult to measure and maintain in the actual surgical operation. In practical cases, the main part of the flexible endoscope is inserted into the patient’s body so that the sheath bending angle information is not readily accessible. In addition, bending angle status could change during surgical operation if patients were to change body position or the endoscopist was to alter the insertion depth of the endoscope. The uncertainty of sheath configuration during the surgical operation poses great difficulties in implementing robust distal-end force prediction models as well as accurate compensation for end-effector position control.

The applications of machine learning (ML) have received increasing attention in the fields such as machinery and robotics [24,25]. Such techniques endow the machine with the ability to gain an intuition of complex systems by processing vast amounts of data. Xu et al. employed a data-driven approach to encoding system inverse kinematics in TSMs based on regression methods such as extreme learning machines, Gaussian mixture regression and K-nearest neighbours regression [26]. In the field of robotic manipulator control, He et al. introduced Iterative Learning Control (ILC) to address the vibration control for an Euler-Bernoulli beam system under aperiodic distributed and boundary disturbance [27,28]. Zhao et al. compared the performances that several state-of-the-art deep learning (DL) algorithms have to offer when applied to monitor machine health using system data collected from low-cost sensors [29]. In these applications, ML is used to efficiently draw insights from the huge data-sets and assist human decision-making. Another popular application of ML is to build predictive models for important system variables such as force. Irgolic et al. applied multi-layer perceptron to make 3-D cutting force prediction of milling functionally graded material by CNC machine [30]. The force prediction has shown reliable results of less than 10% error which allowed CNC machines to operate more safely. Researchers have applied different artificial neural network (ANN) architectures to replace complex mathematical models and express muscle activation force dynamically based on electromyographic (EMG) signals [31–33]. In the case of controlling an under-observed robotic system with uncertain compliant environmental contact, adaptive fuzzy neural network based control provides good robustness and stability [34]. To estimate the distal-end force in a tendon-driven NOTES robot, Li et al. proposed a DL approach that can produce accurate force predictions while identifying the correct force hysteresis profile [35]. However, like many conventional mathematical modeling methods in the literature, this approach works under the assumption of constant and known TSM bending angle. In a nutshell, there are emerging modern ML algorithms that demonstrate remarkable performances in tasks that were conventionally handled by mathematical models.

In this paper, we propose a two-steps data-driven approach to predict the distal-end force of a TSM-driven NOTES robot. Firstly, an ANN model is used to estimate the current TSM cumulative bending angle in the system. Secondly, a combination of two pre-trained ANN models with the nearest bending angles to the estimated result is used to make dynamic estimations of the distal-end tendon force. The ultimate goal of this approach is to provide the distal force information of TSM-driven robots without the demand for additional distal sensors during operations. The force prediction should be accurate, continuous, and robustly identifies the system state in different force hysteresis phases. In addition, by estimating the TSM bending angle information, this method also provides a foundation for potentially using the conventional mathematical approaches with the assumption on known system bending angle in practice. In order to validate the effectiveness of this approach, experiments were conducted on an actual TSM-driven flexible NOTES robot [1].

This paper is organized as follows: an introduction of the properties of TSMs and the experimental set-up will be presented in Section 2. Section 3 explains the detailed methods taken to solve the problem. The results will be analyzed and discussed in Section 4. Finally, we offer our conclusions and directions for future work in Section 5.

2. The flexible endoscopic robotic platform

2.1. Tendon-sheath mechanism

Tendon-sheath mechanisms (TSMs) are commonly used for force transmission in systems with long and tortuous structures such as NOTES and soft wearable robotic systems [36]. As shown in Fig. 1(a), TSM has two main components: an actuation cable (tendon) that transmits force and movement, and a hollow coil (sheath) that wraps around the tendon. TSM offers several advantages including lightweight, compactness, high payload transmission capability and low cost. However, TSM is intrinsically associated with nonlinear friction between the tendon and sheath that causes backlash, hysteresis, and tension loss from the proximal-end force to the distal-end force, which is defined as $f_{TSM} = F_{prox} - F_{dist}$. The relationship between $F_{prox}$ and $F_{dist}$ is often depicted to characterize the unique properties of TSMs, also known as the force hysteresis profile, as illustrated in Fig. 1(b). There are three distinct phases: the pulling phase (phase $\phi$, where the proximal motor winds in and $f_{TSM}$ increases), the releasing phase (phase $\phi$, where the proximal motor unwinds and the tendon relaxes back to its non-tensioned state) and the transition phase (phase $\phi$, where $f_{TSM}$ changes its direction and the system operates in the vicinity of zero velocity). In the transition phase, $F_{prox}$ can change drastically while the $F_{dist}$ remains almost constant, which is commonly referred to as the deadzone. Therefore, it is crucial to identify the correct force hysteresis phase during a force estimation task as $F_{prox}$ and $F_{dist}$ exhibit different relationships in the three phases. In addition, Fig. 1(c) shows how the friction loss changes in different hysteresis phases.

Phee et al. [37] analyzed the mechanics of TSMs by describing the non-linear relationship of $F_{prox}$ and $F_{dist}$ in the three hysteresis phases. Later, Wang et al. [38,39] improved and generalized the model, and the representation is summarized in Eqs. (1-3):

$$F_{dist,\phi} = F_{prox} e^{-\mu \xi}$$

(1)

$$F_{dist,\text{trans}} = F_{dist,\text{pull,last}}$$

(2)

$$F_{prox,\text{trans}} = F_{prox,\text{trans}} e^{-2 \mu \phi}$$

(3)

Where $\xi$ denotes the direction of relative movement of the tendon to its associated sheath and $\mu$ is the friction coefficient between them. $F_{prox,\text{trans}}$, $F_{dist,\text{trans}}$ represent proximal force at the start and end of the $\phi$ transition phase as shown in Fig. 1(b). $\phi$ is the cumulative bending angle in the sheath configuration.
The important conclusion from these studies is that for a given set of tendon and sheath, \( F_{\text{dia}, \phi} \) could exhibit different relationships with \( F_{\text{prox}, \phi} \) depending on the system sheath configuration. More specifically, according to Eqs. (1) and (3), the relationship between \( F_{\text{dia}, \phi} \) and \( F_{\text{prox}, \phi} \) primarily depends on the system cumulative bending angle \( \phi \). Fig. 2(a) presents three different force hysteresis profiles measured on the same NOTES robot when the cumulative bending angles are 100°, 250° and 400°, respectively. Under an identical actuation movement, higher degree bending angle configuration exhibits a slower slope in the pulling phase \( \Theta \) and lower \( F_{\text{dia}, \phi} \) in transition phase \( \Theta \). This is because as the TSM path become more tortuous, there will be an increasing friction loss between the proximal and distal ends.

### 2.2. Experimental setup

This paper proposes an approach to estimating the distal-end force of TSMs for NOTES robots. To ensure the practicality and reproducibility of the experiment, we validated this method by using a flexible endoscopic surgical robot that has passed a live animal study [1]. Fig. 3(a) shows the essential components of the surgical robot used in this study. The robot features a grasper with three Degrees of Freedom (DOF) (yaw, pitch and gripping) driven by three pairs of TSMs, separately. The tendons are connected back to the actuation housing and are driven by DC motors (2657W024CR from Faulhaber Schonhah®) and the proximal movements are measured with encoders (HEDS-S540 A14, 500CPR) attached on each motor including displacement and velocity of the tendon \( \dot{e}_{\text{prox}}, \dot{v}_{\text{prox}} \). The sheaths of the tendons are stopped at the entrance of actuation housing by corresponding load cells (FUTEK LTH300) which are used in this experiment to measure the compression force at the proximal-end of the TSMs. It was proven by Lai et al. [14] that in a TSM, at any given location along the path, the tension in the tendon is equal to the compression force in the sheath. Therefore, the load cells reflect the actual proximal-end tension in each tendon. This robot employs two motors to maintain the bidirectional motion of each DOF. Hence, when one motor (active motor) winds in, its counterpart (passive motor) unwinds to release the tension on the passive tendon. Six load cells are imbedded in the actuation housing to record \( F_{\text{prox},1} \) to \( F_{\text{prox},6} \), respectively, during the operation. We used 0.5 mm diameter tendon together with 1.14 mm outer diameter sheath both from Asahi Intec® for movement transmission. The movements are controlled by PID controllers and all sensory data was acquired and synchronized by MATLAB SIMULINK interfaced with QPIDE data acquisition board from Quanser Quarc®. Sampling rate is set to 1000 Hz for sensor readings and analog output processing, which is sufficient for this experiment and real-world operations. 3D printed barriers are fixed on the supporting acrylic board by screws, which are used to constraint the configuration of TSM. The cumulative bending angle \( \phi \) is calculated from point A (this is where the robotic arm goes into the endoscope during the surgical operation) to point B: end-effector. For instance, as demonstrated in Fig. 3(b), the \( \phi \) between \( \Theta \) and \( \Theta \) is 90°, and the \( \phi \) between \( \Theta \) and \( \Theta \) is 270°. Since the segments in between \( \Theta \) and \( \Theta \) all curve in one direction, their bending angle is cumulated in one go. The path of the TSM is organized in a way that is similar to the route configuration of the gastrointestinal tract.
2.3. Working principle and integration method of the FBG-based sensor

Although the approach proposed in this paper does not require additional sensors during operation, it is crucial to have accurate distal-end force measurements for supervised learning during the training phase. Since the end-effector has an outside diameter of only 4.2 mm, it is difficult to fit in conventional force sensors. Therefore, a Fiber Bragg Gratings (FBG) force sensor was used in this study to measure the distal forces to provide labelled training data for ANNs.

FBGs are generally photo inscribed in the silica core of a single mode optical fiber by using a UV laser beam through a phase mask [40]. FBG reflects a specific wavelength of light called Bragg wavelength \( \lambda_B \), which shifts with respect to the changes in temperature and strain, as shown in the Eq. (4) [41].

\[
\Delta \lambda_B = (K_T \Delta T + K_s \Delta e) \lambda_B
\]  

(4)

where \( \Delta T, \Delta e, K_T, \) and \( K_s \) are the changes in temperature, axial strain along the fiber, thermal sensing coefficient, and strain sensing coefficient of the fiber material, respectively. In this study, to minimize the temperature cross-sensitivity, all the sensing experiments were executed at a constant room temperature of 24.5 °C ± 0.2°C.

The FBG-based force sensor\(^1\) [14] was mounted at the distal-end of a TSM in the robot, as shown in Fig. 4(a-c). The sensor is FBG fiber attached on a nitinol tube welded on a sheath, and it is located at the closing-jaw TSM of the grasper. When closing jaw action is executed, the associated tendon is pulled at the proximal end by the active motor, and simultaneously the sheath experiences compression, which is then transmitted to the force sensor on the sheath. It was proven that at any cross-section (labelled as 1 to 4), the tension T on the tendon and the compression force C on the sheath have a relationship of \( T = -C \). Thus, the FBG sensor captures the strain changes of the nitinol tube due to this compression, and the corresponding wavelength shift reflects the distal-end tension. A prototype of the end-effector with a transparent outer tube is illustrated in Fig. 4(c) to visualize the actual mounting of the force sensor. To fabricate this force sensor, a 1mm-grating FBG sensor (Technica, USA) is bonded with a 3mm-long nitinol tube using epoxy, and this tube is welded in between two sheaths, where the distal sheath has a short length of 3 cm. The force sensor has a sensitivity of 33.09pm/N with a Root-Mean-Square-Error (RMSE) of 8.50pm, which indicates a measurement error of 0.26 N. In this work, wavelength data was collected in real time through an optical sensing interrogator (Micron Optics si255, USA), with a scan rate of 1000 Hz.

\(^1\) A Singapore provisional patent (10201804332S) has been filed on this force sensor.
3. Methodology

3.1. Recurrence plot

Recurrence plot (RP) is a graphical tool devised to compute dynamic parameters from time series. It was initially proposed by Eckmann et al. to visualize the periodic nature of a trajectory through phase space [42]. RP is used for a wide variety of time-series data tasks such as sound analysis, human activity analysis and plant species recognition [43,44]. In other words, RP allows us to convert the multi-dimensional phase space trajectory to a 2-dimensional representation for further analysis purposes. In this study, to reconstruct the phase space, we used autocorrelation function to find a suitable time delay $\tau$. In addition, false nearest neighbours method [45] is applied to estimate the optimal number of embedding dimension $m$. The general recurrence plot is defined by:

$$ R_{ij} = \Theta(\epsilon - \|x_i - x_j\|), \quad x_i \in \mathbb{R}^m, \quad i,j = 1 \ldots N $$

(5)

where $\Theta()$ is the Heaviside function, $\epsilon$ is the threshold distance, $\|\cdot\|$ represents the Euclidean norm, and $m$ is the embedding dimension for $N$ number of considered states $x_i$. In our case, $x$ is the state vectors of load cell time-series data, i.e., the reading of the corresponding load cell. For the robotic gripper, each DOF is driven by two tendons, and each tendon has one load cell at its proximal end. Fig. 5(a) shows the readings of the two load cells of the two tendons controlling the gripping DOF, i.e., the system response, during the probing process (details described in Section 3.4.1). Note that one signal always decreases when the other increases. This is because the active tendon will experience increasing tension while the passive tendon’s tension decreases. Two versions of recurrence plots are constructed for these two particular time series (or load cell readings). Parameter identifications for embedding dimension and time delay are further discussed in Section 4.1. Fig. 5(b) is the unthresholded RPs while Fig. 5(c) is the thresholded RPs with $\epsilon = 0.5$ (we have normalized the array to in between 0 and 1). In this study, as we aim to retain the most texture and color information for the ANN to learn, we have adopted the unthresholded RP approach which is defined by Eq. (6):

$$ R_{ij} = x_i - x_j, \quad x_i \in \mathbb{R}^m, \quad i,j = 1 \ldots N $$

(6)

3.2. Convolutional neural network

In the recent revolution of the deep learning algorithm, convolutional neural network (CNN) is one of the most commonly used architectures that has demonstrated good efficiency in many fields including computer vision [46], natural language processing [47], and pattern recognition based on sensory data [48]. As a supervised learning algorithm, CNN is trained on a labelled dataset and learns a set of weights that minimizes the objective function (the error metrics) specified by the users. It can be built as a regression model that computes a system variable as its output, or as a classification model that produces an output in the form of a vector of categorical scores. Through a series of convolution and pooling operations, CNN is capable of extracting spatial features from different positions of the input which, for most cases, proves to have a more robust performance as compared to manually engineering features for classification tasks. Fig. 6(a) demonstrates the architecture of CNN used in this study. The network takes a pair of RPs as input and computes output in the form of a vector of softmax scores.

3.3. Recurrent neural network

Recurrent neural network (RNN) is a class of ANN that stands out when tackling time-series sequential data. RNN is commonly applied to various research problems such as machine translation [49], sentiment classification [50], and control theory [51] because of its strength in capturing long-term dependencies within time-series data. In the recent development of TSM-driven surgical robotics, Li et al. [35] proposed an approach to taking advantage of an RNN structure called Long-Short Term Memory (LSTM) to make dynamic distal-end force predictions based on proximal sensory data. This method is able to adapt to changing system operation velocities, identify the current hysteresis phase and provide accurate force estimations even when manual manipulation is considered. The recurrent structure of LSTM allows it to extract high-level temporal features between successive time steps while using regulation gates to handle the problem of vanishing gradient. The drawback of this approach, like many other mathematically modeling approaches, is that it was developed based on the assumption of invariant and known shear bending angle in the path. The method proposed in Ref. [35] is incorporated in this study to make distal-end force prediction with a few modifications in its architecture and choice of input sensory data which is shown in Fig. 6(b). The hyperparameters of the LSTM are determined by using Bayesian Optimization with tree-structured Parzen estimators (TPEs) [52].

3.4. Force estimation strategy

In order to make an accurate distal-end force prediction in robotic-assisted NOTES system, we propose a two-stage strategy. In stage one,
the goal is to estimate the current cumulative bending angle of the system through CNN classification and then select two best performing LSTM models, accordingly. In stage two, the selected models are then used to make force predictions during actual system operation.

3.4.1. Stage one

As depicted in the flowchart in Fig. 7(a), the robotic system is initialized with 2N pre-tension on each individual tendon. Next, a probing signal was sent to the gripping DOF of the end-effector. Fig. 7(b) describes the probing signal which is essentially the movement reference for the motors driving the gripping DOF. Various types of probing signals were examined in this study, and we found that a signal with a changing velocity and local maximal amplitude tends to produce a good performance. The probing process lasts for 10s. In the meantime, the proximal forces on the two TSMs of the gripping DOF were recorded by the corresponding proximal load cells as raw data. The data was filtered and normalized, and then constructed into recurrence plots. Next, the recurrence plots are sent to a CNN classifier after being resized to 256 x 256 pixels. The CNN classifier assigns a softmax score, as described in Eq. (7), to each categorical class which represents the probability of the current system cumulative bending angle being a certain angle. The normal range of cumulative bending angle in the NOTES procedure is within 90° to 450°. In this study, we consider 13 classes within this range. The bending angles are selected uniformly with 30° interval, namely, (90°, 120°, 150°, ..., 390°, 420°, 450°). In other words, the classifier receives RPs and computes an output in the form of a vector which contains 13 softmax scores corresponding to the above 13 classes.

\[ P(Y = k|X = x_i) = \frac{e^{s_k}}{\sum_{j=90}^{450} e^{s_j}} \quad \text{where} \ s = f_{\text{CNN}}(x_i; \theta) \]  

Eq. (7) is the standard softmax function used in this study that computes the probability of the current bending angle being k when CNN is given the RP input \( x_i \). \( k = 90°, \ldots, 450° \) and \( \theta \) represents the weights in CNN after training. On the other hand, we have trained 13 LSTM models for each predefined bending angle used for the force prediction in stage two. Two LSTM models with the highest softmax scores are selected correspondingly. The details of LSTM models will be introduced in Section 3.4.2 Stage two.

All training data for this study are collected from the actual flexible NOTES robot. For CNN, we have collected 30 sets of data for each of the 13 classes of bending angles. There are slight state variations for each run of the system due to factors such as initialization and instrumental deviations. We found that 30 sets of data are sufficient for the CNN model to learn and generalize these differences. As introduced in Section 2.2, we have two load cells attached to each DOF; thus the raw data is recorded in two channels \( \{F_{\text{prox1}}, F_{\text{prox2}} \} \) when probing signal is sent. The data is filtered with a second-order low-pass filter and normalized using the mean and variance of each individual channel. After rescaling, the data is used for the CNN training as constructed recurrence
plots in the shape of (256, 256, 2), where 2 represents the number of channels.

In supervised learning, the training dataset needs to be labelled. Instead of labeling the 13 classes using vanilla One-Hot-Encoding (OHE) technique [53], which is the most prevalent strategy for labeling CNN classification training data, we applied a modified version of OHE. By using conventional OHE, we would be assigning 1 to the true class label and 0 for all other classes. Alternatively, we assigned 0.9 to the true class label and diffused 0.05 to its two neighbouring classes. This technique allows the classifier to establish an understanding of the consecutiveness of the 13 classes while still preserve the dominance of the true label. The benefit of doing so is that the CNN model will perceive the bending angles as more of ordinal categorical labels instead of nominal categorical labels, as what we usually do with image classifications. For instance, the bending angle 180° class is labelled as (0, 0, 0.05, 0.05, 0, 0, 0, 0, 0, 0, 0, 0).

3.4.2. Stage two

In stage two, we deploy the two LSTM models selected in the previous stage and use their corresponding softmax scores as weights to make combined predictions of the distal forces of the TMS. During system operation, the proximal sensors on the surgical robot provide inputs for the LSTM models including displacement, velocity, and proximal forces on the two associated tendons. These sensory data are filtered and normalized before being sent to the two LSTM models. Finally, we take a weighted average of the predictions from the two LSTM based on their previous softmax scores, as described in Eq. (8):

\[
LSTM_{final} = \frac{P_{k1} \times LSTM_{k1} + P_{k2} \times LSTM_{k2}}{P_{k1} + P_{k2}}
\]

where \( P_{k1} = P(Y = k_1 | X = x_i) \); \( P_{k2} = P(Y = k_2 | X = x_i) \)

(8)

where \( k_1 \) and \( k_2 \) are the two bending angle classes with the highest scores, while \( LSTM_{k1} \) and \( LSTM_{k2} \) are the model predictions, correspondingly.

The training data for LSTM is collected by proximal sensors in the NOTES system as described in Section 2.2 in the form of \(<F_{\text{prox1}}, F_{\text{prox2}}, F_{\text{prox3}}\>). The label data for training is collected by using the FBG-based force sensor \( F_{\text{dist.larger}} \). For the actuation signal, we adopted the method proposed in Ref. [35] which is a computer generated non-periodical movement in order to ensure the robustness of the trained LSTM models, as shown in Fig. 8(b). Training data volume of 1000 k time steps are used for each model. This process is repeat 13 times to build models for each individual bending angle class. Note that, in this approach, the FBG-based force sensor is only required during the training data collection. During the NOTES robot operation, the distal-end force sensor will be no longer needed.

In this approach, the proposed algorithm has only been trained with the 13 classes of cumulative bending angle that we have predefined. In reality, there are infinite possibilities of cumulative bending angle in the system. However, it is unrealistic to build an infinite number of LSTM force prediction models. Furthermore, it was proven that, in TSMs, the force transmission is solely dependent on the system cumulative bending angle but not how the angle is distributed [23]. Therefore, we choose to build a limited number of LSTM models to cover the cumulative bending angle domain. In this study, we ultimately decide to cover the bending angle domain from 90° to 450° with 13 LSTM models by 30° uniform interval, which is referred to as resolution. As the resolution increases, we will have more accurate predictions. On the other hand, as the demand for resolution goes up, the time required for data collection and model training would also increase consequently. In the short, it’s a trade-off between accuracy and time and training effort. The 30° uniform interval resolution achieves decent force prediction results which will be presented in Section 4.2 Stage two results.

4. Experimental results and discussion

4.1. Stage one results

During the probing process, there are two load cells recording the response of the system’s gripping DOF in the form of proximal-end tendon tension in channel 1 and channel 2, respectively. In order to find suitable embedding parameters for phase space reconstruction, we applied auto-correlation on each channel and identified the time delay \( r \). The optimal \( r \) is identified as the distance between the center peak and the first local minima. The embedding dimension, \( m \), is the number of time delayed copies of the time-series signal used to track its trajectory in phase space. False Nearest Neighbours (FNN) method is applied in this case to determine the minimum number of embedding dimension required for these two channels as shown in Fig. 9. Therefore, we have \( \{ r = 85, m = 4 \} \) for channel 1 and \( \{ r = 55, m = 5 \} \) for channel 2 that minimize the average mutual information function [54] during the reconstruction of the recurrence plots in their corresponding channels. Note that channel 1 takes longer time delay for reconstruction and these embedding parameters are specific to the chosen probing signal.

Fig. 10 visualizes the RPs that are sampled from different bending angles from the two channels. RPs from the same channel depict very similar texture due to the fact that they share the same actuation signal during the probing procedure. Therefore, they would have similar behaviours in their reconstructed phase space. Moreover, the intensity of the sub-regions demonstrates more distinct gradual differences across bending angle classes which allow the CNN model to effectively perform classification.

As mentioned in the CNN training strategy, 30 training datasets are collected for each bending angle class. To evaluate the performance of the CNN model, we applied the five-fold cross-validation (CV) protocol. The mean accuracy of the CV is 98.72%. Finally, the CNN model is trained with all training data.

In order to validate the effectiveness of the proposed approach, we evaluated 4 different bending angles: 170°, 225°, 315°, and 400°. Table 1 shows the CNN prediction results:

- We tested bending angles that are not exactly the same with the training bending angles as this is what the algorithm is likely to encounter in the real case scenarios. By the definition of softmax score, the sum of all scores across all bending angle classes is equal to 1. We observed from Table 1 that, for each individual test set, the CNN model has assigned most of the score to the adjacent bending angle classes to the test bending angle (over 90%). In addition, the closer the test bending angle is to the training bending angle classes the higher the scores. During the actual surgical operation, when the test bending angle is actually unknown, this prediction would provide us with a reasonable estimation of the bending angle. More importantly, the prediction results allow us to select two winning models, namely \( LSTM_{k1} \) and \( LSTM_{k2} \), and softmax scores that are then used in stage two to compute the weighted average for force prediction.

4.2. Stage two results

Fig. 11 shows the stage two distal force prediction results for cumulative bending angles 170°, 225°, 315°, and 400°, respectively. As we can observe from plots (a)-(d), by using an LSTM model trained with a smaller bending angle than the testing bending angle would result in over-predictions as compared to ground-truth FBG readings, whereas by using an LSTM model trained with larger bending angle produces under-predicted forces. This tendency accords with the TSM characteristics as described in Fig. 2, i.e., with smaller bending angles there is smaller system friction; thus, the distal-end force is expected to be larger given the
Fig. 8. (a) A flowchart of the Stage two pipeline. Distal forces are estimated by the two LSTM models (selected in stage one) based on the proximal sensory readings of the robotic system during operation. (b) A sample of 300 s of non-periodical movement reference for training data collection.

Fig. 9. (a) In order to construct RPs, auto-correlation is used to determine time delay, \( \tau \). (b) FNN is used to determine embedding dimension, \( m \).

Table 1

<table>
<thead>
<tr>
<th>CNN bending angle predictions.</th>
<th>( k_1 ) bending angle</th>
<th>( k_1 ) softmax score</th>
<th>( k_2 ) bending angle</th>
<th>( k_2 ) softmax score</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \psi ): 170° 150°</td>
<td>0.3440</td>
<td>180°</td>
<td>0.6162</td>
<td></td>
</tr>
<tr>
<td>( \psi ): 225° 210°</td>
<td>0.4269</td>
<td>240°</td>
<td>0.5655</td>
<td></td>
</tr>
<tr>
<td>( \psi ): 315° 300°</td>
<td>0.5335</td>
<td>330°</td>
<td>0.4458</td>
<td></td>
</tr>
<tr>
<td>( \psi ): 400° 390°</td>
<td>0.6394</td>
<td>420°</td>
<td>0.2734</td>
<td></td>
</tr>
</tbody>
</table>
same proximal-end force, and vice versa. Therefore, the actual distal-end force falls in between the two selected LSTM models’ predictions. Plots (c)-(h) demonstrate that the algorithm can track the distal force in different hysteresis phases and the error is relatively higher in the transition phase. A potential solution is discussed in future work.

In this approach, we use the models’ softmax scores to compute the final distal-end force prediction, which is plotted in red in Fig. 11. From Table 2, we can see that the weighted average prediction demonstrates a better estimation than by using merely one adjacent bending angle LSTM model, namely $k_1$ and $k_2$ bending angles. The mean RMSE for the proposed approach is 0.1197 N. Moreover, we have observed that the algorithm makes more accurate predictions in the $\psi$ pulling and $\varphi$ releasing phases than in the $\varphi$ transition phase. This observation also corroborates with the learned force hysteresis profiles of different testing bending angles, as shown in Fig. 11(e)-(f). The X-axis and Y-axis interval are maintained at a 2:1 ratio to better visualize the shape of different hysteresis profiles. The algorithm can robustly identify the three hysteresis phases, more importantly, it adapts to the changing slope in phase $\varphi$ and $\psi$ across different cumulative bending angles, both of which are valuable traits that conventional mathematical modeling methods do not bear. However, the distal-end force prediction in phase $\psi$ demonstrates more errors such as under-prediction, over-prediction, and fluctuation. The maximum error occurred in these predictions ranges from 0.3 N to 0.5 N. There are two main reasons for this phenomenon. Firstly, the distal-end force in phase $\psi$ is affected greatly when considering different cumulative bending angles. Secondly, when the FBG-based force sensor is used to measure the distal-end force for the training data labeling, the sensor is essentially measuring the compression force on the associated sheath. As a result, when the system friction changes direction in transition phase $\varphi$, there is a micro-movement observed in the sheath and sometimes unexpected contact with its neighbouring sheaths. Therefore, the distal-end force measured in this particular phase is more susceptible to the system disturbance as compared to other approaches that feature a load cell fixed on a stand, which is more prevalent in the literatures [21–23,35]. This is also one of the key challenges of applying FBG-based force sensor directly on the NOTES robot end-effector as it may pick up unforeseen wavelength shift during complicated surgical procedures which does not reflect the actual tension change in the tendon itself.

The time taken to collect all the training data for this approach is around 26 h. The total training time for CNN and LSTM models is 11.4 h. Our PC configuration is as follows: Intel® Xeon® CPU E5-1620 v2 @ 3.70GHz, RAM-16.0 GB, and NVIDIA GeForce GTX 1080 Ti. In stage one, the probing process itself lasts for 10 s after initializing the system and the times took to process an RP and make a prediction with the CNN model are 75.7 ms and 4.54 ms, respectively. In stage two, the average time needed for force prediction for one time step is 0.311 ms, which is sufficient for the 1000 Hz use case.

4.3. Classification vs regression

As discussed in Section 3.5, besides using diffused OHE technique, we have also considered applying vanilla OHE labeling and building the CNN as a regression model. For the CNN regression model, the training data are directly labelled by their corresponding cumulative bending angles: (90°, 120°, ... 420°, 450°). Consequently, when given a testing RP, a prediction is given in the form of an estimated bending angle. Then two models are selected from the LSTM pool which are adjacent to the predicted bending angle. E.g. the CNN model outputs 100°, then

---

**Table 2**

<table>
<thead>
<tr>
<th>$\varphi$ (deg)</th>
<th>$k_1$ bending angle RMSE (N)</th>
<th>$k_2$ bending angle RMSE (N)</th>
<th>Weighted average RMSE (N)</th>
<th>Mean RMSE (N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>170°</td>
<td>0.2872</td>
<td>0.1854</td>
<td>0.0918</td>
<td>0.1197</td>
</tr>
<tr>
<td>225°</td>
<td>0.3077</td>
<td>0.3528</td>
<td>0.1095</td>
<td></td>
</tr>
<tr>
<td>315°</td>
<td>0.3149</td>
<td>0.5098</td>
<td>0.1008</td>
<td></td>
</tr>
<tr>
<td>400°</td>
<td>0.2061</td>
<td>0.2920</td>
<td>0.1767</td>
<td></td>
</tr>
</tbody>
</table>

---

**Fig. 10.** (a) RPs sampled from load cell channel 1 across different cumulative bending angles. (b) RPs sampled from load cell channel 2 across different cumulative bending angles.
\( \text{LSTM}_{90^\circ} \) and \( \text{LSTM}_{120^\circ} \) will be selected. The force prediction is computed by using linear interpolation.

Tables 3–4 show the experimental results by using vanilla OHE and regression technique. The same testing dataset was used in this case, namely, \( \{170^\circ, 225^\circ, 315^\circ, 400^\circ\} \). The goal of stage one bending angle estimation is to select the best performing LSTM models and compute weighting for the second stage. As compared to diffused OHE model, although the vanilla OHE model selects the same LSTM force prediction models, the algorithm showed a tendency to overly dependent on one of the training bending angle classes which is commonly referred to as to produce ‘harder’ predictions. As a result, the mean RMSE goes up to 0.1788 N for the vanilla OHE CNN model. The regression approach suffers from the similar problem and has a mean RMSE of 0.2039 N. Therefore, the CNN classification model with diffused OHE is adopted for this approach as it produces a ‘softer’ score when the testing bending angle falls in between two training bending angles. The accuracy comparison of the three techniques are shown in Fig. 12.

4.4. Robustness against bending angle distribution

Robot-assisted NOTES are mostly used to perform Endoscopic Submucosal Dissection (ESD) which is an advanced surgical procedure used to remove early-stage gastrointestinal tumours in the stomach or colon [55]. When the endoscope delivers the robotic arms to the patient’s
stomach, through the esophagus, the path configuration is relatively predictable, as shown in Fig. 13, path 1. On the other hand, when ESD is conducted in the patient’s colon, the endoscope enters via the rectum and reaches the surgical site through the intestinal tract. The configuration of the large intestine does not vary drastically from patient to patient, however, the bending angle distribution could be slightly different depending on patients’ size, bodily position, and individual difference.

According to Eqs. (1–3), the force hysteresis profile is essentially determined by the cumulative bending angle $\psi$ in the TSM. In addition, it was proven by a series of experiments [23] that when given the same cumulative bending angle but different bending angle distribution, the TSM demonstrates identical friction force profiles. Moreover, to further validate the algorithm’s robustness in this regard, we tested the TSM configurations of a fixed cumulative bending angle with three different bending angle distributions. Fig. 14 depicts a cumulative bending angle of 360$^\circ$ with three different bending angle distributions, where (a) simulates patients with a bigger size, (b) simulates patients with a smaller size, and (c) simulates the cases where there are sharper angles in between the transverse colon and the descending colon. Configuration (a) is used when collecting training data for LSTM$^{\text{S}0}$, (b) and (c) are referred to as test_config_1 and test_config_2, respectively. Note that for this case we take 3 LSTM models for force prediction because that the summed softmax score of $k_1$ and $k_2$ is lower than 0.9 which only happens when the testing cumulative bending angle is very close to one of the training bending angle. The result RMSEs are 0.1448 N and 0.1487 N, respectively. We found that the RMSE is slightly higher when considering various bending angle distributions. This is verified with another two cases when the cumulative bending angles are 280$^\circ$ and 410$^\circ$ where we set-up the testing angle distributions to be dissimilar to the training angle distributions. As shown in Table 5, the mean RMSE is 0.1711 N which is sufficient for the NOTES surgical use, and this is a more practical error estimation.

**Table 3**
Vanilla OHE CNN results.

<table>
<thead>
<tr>
<th>$\psi$</th>
<th>$k_1$ bending angle: score</th>
<th>$k_2$ bending angle: score</th>
<th>RMSE</th>
<th>Mean RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>170$^\circ$</td>
<td>150$^\circ$: 0.0003</td>
<td>180$^\circ$: 0.9988</td>
<td>0.0863 N</td>
<td>0.1788</td>
</tr>
<tr>
<td>225$^\circ$</td>
<td>210$^\circ$: 0.2521</td>
<td>240$^\circ$: 0.7384</td>
<td>0.1993 N</td>
<td>N</td>
</tr>
<tr>
<td>315$^\circ$</td>
<td>300$^\circ$: 0.7652</td>
<td>330$^\circ$: 0.2335</td>
<td>0.3167 N</td>
<td>N</td>
</tr>
<tr>
<td>400$^\circ$</td>
<td>390$^\circ$: 0.0243</td>
<td>420$^\circ$: 0.9616</td>
<td>0.2854 N</td>
<td>N</td>
</tr>
</tbody>
</table>

**Table 4**
Regression CNN results.

<table>
<thead>
<tr>
<th>Model prediction</th>
<th>$k_1$: weightage</th>
<th>$k_2$: weightage</th>
<th>RMSE</th>
<th>Mean RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\psi$: 170$^\circ$</td>
<td>151.16$^\circ$</td>
<td>150$^\circ$: 0.9613</td>
<td>180$^\circ$: 0.0387</td>
<td>0.0786 N</td>
</tr>
<tr>
<td>$\psi$: 225$^\circ$</td>
<td>230.50$^\circ$</td>
<td>210$^\circ$: 0.3167</td>
<td>240$^\circ$: 0.6833</td>
<td>0.1269 N</td>
</tr>
<tr>
<td>$\psi$: 315$^\circ$</td>
<td>306.42$^\circ$</td>
<td>300$^\circ$: 0.7860</td>
<td>330$^\circ$: 0.2140</td>
<td>0.3395 N</td>
</tr>
<tr>
<td>$\psi$: 400$^\circ$</td>
<td>392.46$^\circ$</td>
<td>390$^\circ$: 0.9180</td>
<td>420$^\circ$: 0.0820</td>
<td>0.2704 N</td>
</tr>
</tbody>
</table>
5. Conclusion

Tendon-sheath mechanisms are popularly used as force transmission systems in surgical robotic applications when the size constraint is a priority. However, the non-linear friction between the tendon and its associated sheath causes tension loss, backlash and it is difficult to model. In this study, we propose an approach that addresses common limitations of conventional mathematical modeling methods such as discontinuity, assumptions on no system slack, and most importantly, unrealistic assumptions of known system cumulative bending angles. This study proposes an approach that firstly estimates the current system bending angle by processing an RP constructed from the sensory feedbacks during a probing process. Then, two pre-trained LSTM models are selected to make dynamic distal-end force prediction during system operation. Instead of validating this algorithm on a simulated platform set-up, we adopted the state-of-the-art FBG-based force sensor so that tendon tension could be directly measured on the actual surgical robot in order to ensure the practicality of the proposed approach. The algorithm can robustly adapt to different force hysteresis profiles and accurately identify hysteresis phases. In addition, the cumulative bending angle estimation obtained from this approach could also facilitate the conventional modeling methods that require known bending angle information to be incorporated in practice.

Future activities will be focused on two main areas. Firstly, further improving the precision of the force prediction, especially in the transition phase as it usually contains the local maximal force. This includes designing better ANN architectures and collecting more accurate training data by enhancing the FBG-based force sensor’s performance via means such as isolating the sensor from unexpected surrounding contacts. In addition, we will explore the possibility of constructing end-to-end prediction models and designing automated ways to collect large amounts of data that will be needed for training such models.

Acknowledgements

This research is supported by the National Research Foundation (NRF) Singapore (NRF2016-07).

References


