2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 14-18, 2024. Abu Dhabi, UAE

Control-Oriented Reinforcement Active Modeling Scheme for Hysteresis Compensation of Flexible Endoscopic Robot

Fan Ren, Xiangyu Wang*, *Member, IEEE*, Yongchun Fang, *Senior Member, IEEE*, Yanding Qin, Hongpeng Wang, Ningbo Yu, and Jianda Han, *Member, IEEE*

***Abstract*— Hysteresis has posed significant challenges to the modeling and control of flexible endoscopic robots, which impedes the advancement of automated endoscopic operation. Despite numerous hysteresis modeling approaches aimed at improving accuracy, there are still several unresolved issues, such as inappropriate model selection and non-ideal assumption of noise. Focusing on these challenges, a novel reinforcement active modeling (RAM) scheme is proposed in this paper. By incorporating reinforcement learning, this method augments an Extended Kalman Filter (EKF)-based active modeling strategy, which improves the insensitivity and generalization ability to non-Gaussian noise that is not introduced in training. Finally, a series of comparative experiments are conducted on the self-built flexible endoscopic robot to validate the improvement achieved by the proposed scheme. Compared with some widely-applied methods, the proposed scheme achieved at least 63.8% improvement in the root mean square error (RMSE) in modeling accuracy under Gaussian noise conditions, and at least 36.5% improvement in RMSE under Poisson noise conditions.**

I. INTRODUCTION

Natural orifice transluminal endoscopic surgery (NOTES) has emerged as a prominent technology in minimally invasive treatments, offering unique benefits, such as reduced pain, minimal trauma, and rapid recovery [1]–[4]. The flexible endoscope finds extensive application in NOTES procedures, such as renal calculus ablation operations, due to its softness, non-invasiveness, and wide extension capabilities. However, the inherent flexibility of the tendon-sheath mechanism results in noticeable hysteresis in operating the flexible endoscope, which significantly complicates precise control of the bending motion. While the phenomenological model is initially employed as the hysteresis model to depict input-output mappings, it struggles to accurately describe time-varying uncertainties due to its static features [5]–[7].

This work was supported by National Key R&D Program of China (Grant No. 2022YFB4702800), National Natural Science Foundation of China (Grant No. 62303248), Guangdong Basic and Applied Basic Research Foundation (Grant Nos. 2024A1515010102, 2023A1515110678), China Postdoctoral Science Foundation Funded Project (Grant No. 2023M731804), Shenzhen Science and Technology Program (Grant No. KQTD20210811090143060) and Sustainable Development Science and Technology Special Project of Shenzhen (Grant KCXFZ20230731100900002).

*Xiangyu Wang is the corresponding author.

The authors are with the Institute of Robotics and Automatic Information System, Nankai University, Tianjin 300350, China and also with the Institute of Intelligence Technology and Robotic Systems, Shenzhen Research Institute of Nankai University, Shenzhen 518083, China. (e-mail: renfan@mail.nankai.edu.cn, wangxyu@nankai.edu.cn, fangyc@nankai.edu.cn, qinyd@nankai.edu.cn, hpwang@nankai.edu.cn, nyu@nankai.edu.cn, hanjianda@nankai.edu.cn)

A dynamic modeling scheme, known as active modeling (AM), has been adept at estimating the uncertainties of various kinds of nonlinear systems, including mobile robots [8], unmanned helicopters [9], unmanned aerial vehicles [10], unmanned surface vehicles [11], pneumatic artificial muscles [12], shape memory ally [13], etc. In particular, the active modeling scheme provides an alternative approach for the estimation of the hysteresis effect. For instance, aiming at the limitations of Kalman filter (KF)-based active modeling when applying it to flexible endoscopic robot, Wang *et al.* proposed an unscented Kalman filter (UKF)-based active modeling scheme in which the Coleman-Hodgdon (C-H) model is chosen as reference model, and achieved accurate dynamic modeling for the bending motion of flexible endoscopic robot [14]. With respect to this concept, Qin *et al.* have further extended this nonlinear dynamic modeling scheme, culminating in the amalgamation of the modified methodology with the Bouc-Wen model to mitigate hysteresis nonlinearity in pneumatic artificial muscles, thereby reducing tracking errors [15]. Nevertheless, due to the priori assumption of Gaussian noise and the mismatch in the dynamic characteristics of the reference model, the applicability of these classical active modeling approaches may be limited.

To overcome the aforementioned constraints, an assumption of the confidence interval-based active modeling scheme has been proposed for the estimation of hysteresis boundaries. For instance, Wang *et al.* proposed an extended set-member filter (ESMF) observer-based active modeling scheme to estimate the hysteresis boundaries in flexible endoscopic robots, thereby mitigating modeling errors [16]. Furthermore, Qin *et al.* have proposed an attain modeling scheme founded upon the Prandtl-Ishlinskii (PI) model and ESMF to achieve compensation for system nonlinearity and external disturbances of pneumatic artificial muscle [17]. While ESMF-based approaches assume noise only to be unknown but bounded, ESMF-based methods struggle to achieve precise point estimations within confidence intervals. Meanwhile, the judicious selection of an appropriate reference model remains a vital challenge of active modeling-based schemes.

Data-driven algorithms and intelligent algorithms are introduced to emancipate active modeling from the reliance on reference models in this case. For instance, Chen *et al.* has proposed an innovative data-driven active modeling approach that amalgamates the Koopman operator with Kalman filtering, thereby mitigating the influence of modeling uncertainties and external disturbances on the pneumatic soft

manipulator, enabling accuracy control [18]. It should be mentioned that Koopman-based active modeling remains to be constrained by the assumptions of linear systems and the Gaussian distribution noise. While current learning-based modeling approaches also exhibit rewarding performance, it is acknowledged that exclusively relying on learning for modeling is time-consuming and lower productive [19], [20]. Consequently, challenges remain in finding a reasonable integration of data-driven algorithms into the modeling process.

This paper presents a novel method called reinforcement active modeling (RAM) scheme for accurate hysteresis modeling of the flexible endoscopic robot in the presence of non-Gaussian noise. Specifically, traditional KF-based active models are difficult to capture hysteresis loops in inaccurate initial estimates and non-Gaussian noise, while ESMF-based models fail to provide precise point estimates. To address these limitations, we propose an active modeling approach that integrates reinforcement learning (RL)-based compensation. This method achieves precise state vector estimation in non-Gaussian noise by compensating for modeling errors. The principal contributions of this manuscript can be summarized as follows:

1) A novel hysteresis model that empowers EKF-based active modeling using reinforcement learning is proposed in this paper. This scheme offers a more versatile means to attain accurate modeling, even in the inaccurate initial estimation of the reference model.

2) This paper represents a *pioneering* application of reinforcement learning in the hysteresis modeling for the flexible endoscopic robot. Particularly, higher modeling accuracy is achieved when noise assumptions are non-ideal.

3) The effectiveness of the proposed method is verified through comparison experiments with the C-H model, AM, and the long short-term memory (LSTM) modeling scheme, respectively.

The remainder of this paper is organized as follows: Section II provides an elaborate exposition of the active modeling approach employing the C-H model as the reference model and the EKF as the state estimator. In Section III, the reinforcement active modeling scheme is built up. Section IV offers a detailed description analysis of the experimental results pertaining to the proposed method. Finally, the concluding section provides the research's overall conclusion.

II. Preliminary of Active Modeling

A. Coleman-Hodgdon Hysteresis Model

The Coleman-Hodgdon (C-H) model is a classical model used in describing the phenomenon of hysteresis [21]. In this article, the continuous C-H model is utilized to capture the hysteresis of the flexible endoscopic robot as the reference model [14], which is written as follows:

$$\dot{\theta}(t) = a\left|v_l(t)\right| \cdot (b \cdot \delta l(t) - \theta(t)) + c \cdot v_l(t) \tag{1}$$

where $\theta(t)$, $\delta l(t)$, $v_l(t)$ represent the bending angle, pulling distance, and the velocity of pulling distance, respectively. $v_l(t) = \dot{\delta l}(t)$, and a, b, c are constants. As the schematic of flexible endoscopic illustrated in Fig. 1, the pulling distance

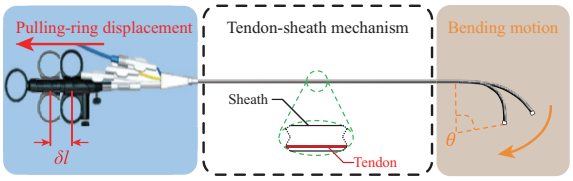

Fig. 1. The schematic diagram of the flexible endoscope's bending motion.

$\delta l(t)$ induces variations in the bending angle $\theta(t)$ of the flexible endoscope. The C-H model provides an approximate description of the mapping relationship between input and output. It is evident that equation (1) remains static once the identification parameters a, b, c are determined. Whereas, this static model is inadequate for handling time-varying uncertainties, which results in unsatisfactory modeling accuracy.

B. EKF-based Active modeling

This paper takes EKF as an example to introduce active modeling. EKF is utilized as the extended state estimation engine to perform real-time estimation of both system states and modeling errors introduced in equation (1) for the reference model. As discussed in previous works [22], this kind of approach treats uncertainties as additive process noise in the following form:

$$\begin{cases} \dot{\theta}(t) = a\left|v_l(t)\right|(b \cdot \delta l(t) - \theta(t)) + c \cdot v_l(t) + W(t) \\ y(t) = \theta(t) + V(t) \end{cases} \tag{2}$$

where $y(t)$ is the measurement of the output equation, $W(t)$ and $V(t)$ represent the process noise and measurement noise respectively, which are both assumed to be Gaussian white noise in real application. The modeling errors of equation (2) can be re-expressed as follows:

$$\begin{cases} f(t) = \dot{\tilde{\theta}}(t) - \dot{\theta}(t) \\ \dot{f}(t) = \vec{0} + \varepsilon(t) \end{cases} \tag{3}$$

where $\tilde{\theta}$ and θ represent the measurement of the bending angle and the output of the C-H model respectively, $f(t)$ is the model error, and $\varepsilon(t)$ is the process noise which drives the model error update. Considering the fact that the updating rate of model errors is much slower than the sampling frequency of the system, the lumped term $f(t)$ can be regarded as a slow time-varying value [12]. Hence, the discrete form of the original model can be re-written as

$$\begin{cases} \tilde{\theta}_{k+1} = \tilde{\theta}_k + \left[a\left|v_{l(k)}\right|(b \cdot \delta l_k - \tilde{\theta}_k) + c \cdot v_{l(k)} + f_k\right] T + W_k \\ f_{k+1} = f_k + \varepsilon_k \\ \tilde{y}_k = \tilde{\theta}_k + V_k \end{cases} \tag{4}$$

where T is the sampling time. Symbols $\tilde{\theta}_k$, δl_k, f_k, W_k, ε_k, V_k, $v_{l(k)}$ represent the discrete form of $\tilde{\theta}(t)$, $\delta l(t)$,$f(t)$, $W(t)$, $\varepsilon(t)$, $V(t)$, $v_l(t)$, respectively; $v_{l(k)} = (\delta l_k - \delta l_{k-1})/T$. The discrete form of state vector can be expressed as $X_k = \begin{bmatrix} \tilde{\theta}_k & f_k \end{bmatrix}^{\mathrm{T}}$. To facilitate the derivation process, equation (4) is rewritten as follows:

$$\begin{cases} X_k = F(X_{k-1}) + W_k \\ Y_k = H(X_k) + V_k \end{cases} \tag{5}$$

where $F(\cdot)$ and $H(\cdot)$ represent the state transition function and the measurement function respectively. Then, the EKF is chosen as the estimator of active modeling to estimate the state vectors in real-time, which is implemented as follows:

Step 1. The initial state and covariance matrix are initialized. Then, the prediction of the state vector and measurement are expressed as follows:

$$\begin{cases} X_{k|k-1}(1) = X_{k-1}(1) + \Big[a \cdot \big|v_{l(k)}\big| \, (b \cdot \delta l_k - X_{k-1}(1)) \\ \qquad\qquad + c \cdot v_{l(k)} + X_{k-1}(2)\Big] \cdot T \\ X_{k|k-1}(2) = X_{k-1}(2) \\ Y_{k|k-1} = X_{k|k-1}(1) \end{cases} \tag{6}$$

where $X_{k|k-1}$ and $Y_{k|k-1}$ represent the prediction of the state vector and measurement at k-th time step, respectively.

Step 2. The state transition matrix and measurement matrix are obtained by linearising the process model and measurement model in equation (5), respectively.

$$\begin{cases} \Sigma_k = \dfrac{\partial F}{\partial X} = \begin{bmatrix} 1 - T \cdot a \cdot \big|v_{l(k)}\big| & T \\ 0 & 1 \end{bmatrix} \\ \Omega_k = \dfrac{\partial H}{\partial X} = \begin{bmatrix} 1 & 0 \end{bmatrix} \end{cases} \tag{7}$$

where Σ_k and Ω_k represent the state transition matrix and measurement matrix at k-th time step respectively.

Step 3. Next, the prediction of the covariance matrix and update of the Kalman gain is obtained as follows:

$$\begin{cases} P_{k|k-1} = \Sigma_k P_{k-1|k-1} {\Sigma_k}^{\mathrm{T}} + Q_k \\ K_k = P_{k|k-1} {\Omega_k}^{\mathrm{T}} (\Omega_k P_{k|k-1} {\Omega_k}^{\mathrm{T}} + R_k)^{-1} \end{cases} \tag{8}$$

where $P_{k|k-1}$ is the prediction of the covariance matrix at k-th time step, $P_{k-1|k-1}$ is the covariance matrix at $(k-1)$-th time step, $\boldsymbol{Q_k}$ and $\boldsymbol{R_k}$ represent the covariance matrix of process noise and measurement noise at k-th time step respectively, and K_k is the Kalman gain at k-th time step.

Step 4. Eventually, the state vector and measurement are updated as follows:

$$\begin{cases} X_k = X_{k|k-1} + K_k(Y_k - Y_{k|k-1}) \\ P_{k|k} = P_{k|k-1} - P_{k|k-1} K_k \Omega_k \end{cases} \tag{9}$$

where X_k and Y_k represent the state vector and the measurement at k-th time step respectively. Therefore, the state vector estimation can be obtained through active modeling. However, it is worth mentioning that despite the improved accuracy of the estimation compared to the output of the C-H model alone, there are still some open issues as follows:

1) The EKF critically depends on the careful selection of the reference model and the appropriate adjustment of its initial state.

2) The estimation accuracy of the traditional EKF is influenced by the initial value chosen for the noise variance.

3) The requirement of Gaussian distribution for noises limits the ability of active modeling schemes to be extended to various types of systems.

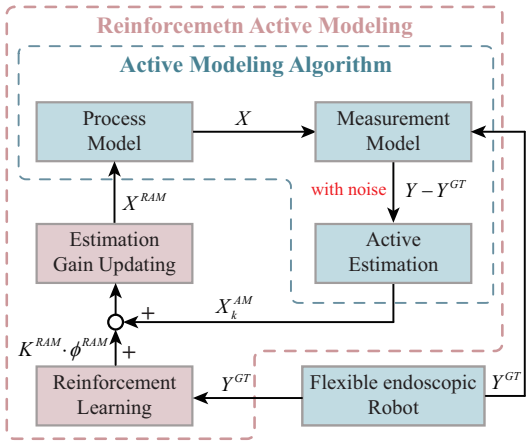

Fig. 2. The structure of RAM scheme.

III. REINFORCEMENT ACTIVE MODELING

In this section, the outstanding approximation and generalization capabilities of reinforcement learning are employed to address the issues mentioned in the last section. Firstly, the interaction between the model and the environment is represented as a Markov Decision Process (MDP). This MDP is denoted as a tuple $< S, A, P, R >$, where S and A represent the set of the states and actions respectively, $P : S \times A \times S \to \mathbb{R}$ denotes the transition probability, and R represents the reward function. In this paper, the state vector $s_k \in S$ represents the discrepancy, denoted as ϕ_k^{RAM}, between the ground truth vector y_k^{RAM} and the output of RAM vector y_k^{GT} at the time step k. The action space vector is the reinforcement gain K_k^{RAM} at the time step k.

It is worth mentioning that the proposed scheme achieves an improvement in active modeling, accordingly, the workflow is designed on the basis of the preceding section. As the structure of the RAM scheme shown in Fig. 2, reinforcement learning is introduced to learn from the residual modeling error between active modeling and ground truth value to reinforcement gain K_k^{RAM}. Subsequently, the residual modeling error is offset via a reinforcement measurement update, obtaining the reinforcement gain from the reinforcement learning policy. The estimation of AM and RAM are denoted by $X_{k|k}^{AM}$ and $X_{k|k}^{RAM}$ respectively. Then, the proposed scheme is expressed as follows:

$$\begin{cases} X_{k|k}^{RAM} = X_{k|k}^{AM} + K_k^{RAM} \cdot \phi_k^{RAM} \\ \phi_k^{RAM} = Y_k^{GT} - Y_k^{RAM} \end{cases} \tag{10}$$

where K_k^{RAM} is the reinforcement gain generated by reinforcement learning, which is used in reinforcement measurement update after active modeling, Y_k^{GT} and Y_k^{RAM} represent the ground truth and the output of RAM, respectively. ϕ_k^{RAM} is the residual error between the Y_k^{GT} and Y_k^{RAM}. Consequently, the reward function R_{kk} measures the effectiveness of action by quantifying the modeling deviation between the ground truth value and the estimation of RAM, i.e. ϕ_k^{RAM}.

$$R_{kk} = -\left\|\phi_k^{RAM}\right\|^2 \tag{11}$$

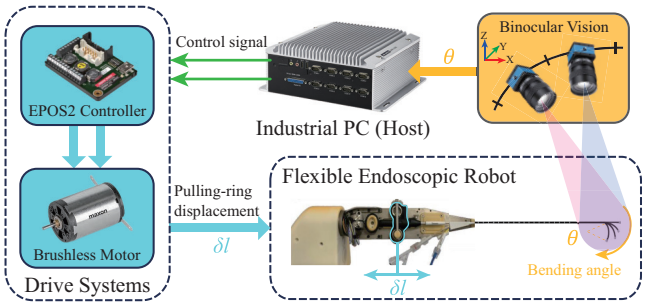

Fig. 3. The block diagram of flexible endoscopic robot system.

TABLE I
HYPARAMETERS OF PPO TRAINING PROCESS

Number	Hyparameters	Value
1	Learning rate	$2.5e-4$
2	Experience horizon	1024
3	Discount factor	0.99
4	Minibatchsize	128
5	Clip factor	0.2
6	Episodes	2000

It is worth noting that the proposed algorithm does not disrupt the integrity of the active modeling. The final state is obtained through dual steps, i.e., active modeling updates in (9) and reinforcement learning updates in (10). The combination of reinforcement learning for residual compensation and active modeling estimation governs the state updates.

IV. Ablation Studies and Analysis

A. Platform Description

As shown in Fig. 3, the platform is composed of a flexible endoscopic robot with a ureteroscope, a binocular vision system, an industrial PC, a driver, and a brushless DC servo motor. The binocular vision system that comprises two industrial cameras (The ImagingSource GigE DFK33GP1300) is used to capture bending angles. The industrial PC transmits motion signals to control the DC servo motor's movement via the driver (EPOS2 24/2 Controller), while simultaneously capturing the bending angle in real-time from the cameras. Numerous experiments are conducted within a MATLAB/Simulink environment.

B. Validation of Modeling Accuracy

In this section, three groups of comparative experiments are carried out to validate the modeling accuracy, fluctuation, and generalization capabilities of the proposed modeling scheme under Gaussian noise and non-Gaussian noise. Direct RL modeling is selected as a comparative scheme to substantiate the distinctive advantage of combining the rule-based active model with the learning strategy. Furthermore, the C-H model, AM, and LSTM modeling schemes have been chosen as representatives of the static, active, and learning-based models, respectively, for comparison with RAM. Utilizing experimental data obtained from the hardware device, a genetic algorithm is employed to identify the parameters of the C-H model, resulting in the following parameter values:

$$a = 1.15,\ b = 22.38,\ c = 5 \tag{12}$$

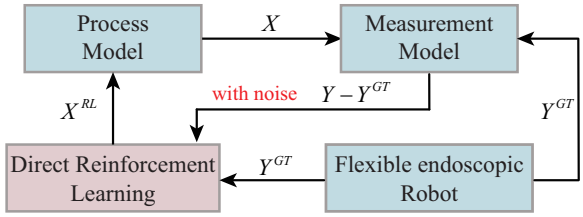

Fig. 4. The structure of Direct RL modeling scheme.

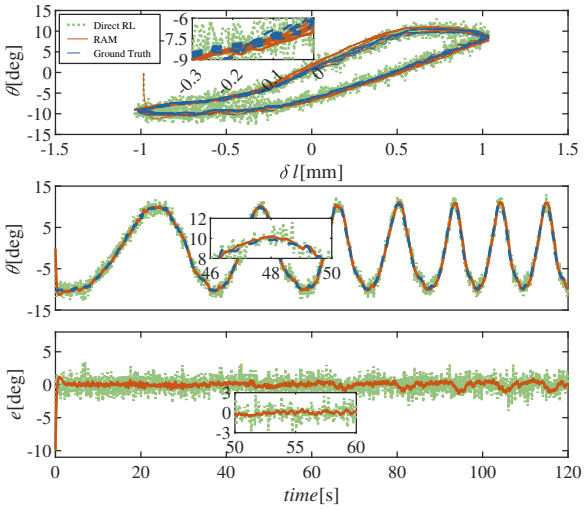

Fig. 5. The results in **Experiment 1** (The ground truth values-blue dashed line; the RAM-orange solid line; the Direct RL modeling-green dotted line).

It is worth mentioning that the corresponding dataset in pulling distance $\delta l \in [2, 4]$ mm is utilized as the training set. During training, the measurement noise is selected as $\boldsymbol{R_k} = 0.1$. The dataset in pulling distance $\delta l \in [-1, 1]$ mm is selected as the testing set. During testing, the covariance of Gaussian noise which is not introduced in training is set as 1 and 5 respectively. The expectation of Poisson noise is set as 0.1, 1, and 2. In order to ensure the validity of the experimental results, RAM and Direct RL modeling applied the Proximal Policy Optimization (PPO) algorithm as the RL algorithm. The hyperparameters for the PPO training process are set to identical values as displayed in TABLE I. Additionally, the training datasets, testing datasets, and other RL settings of RAM are also set to be the same as Direct RL modeling.

Experiment 1 *(comparative experiments of Direct RL modeling and RAM):* To validate the necessity of the proposed modeling scheme, an experiment is conducted to highlight its superiority and compare the modeling accuracy and fluctuation of Direct RL modeling and RAM. It should be noted that in this particular case, RL is used as a substitute for active modeling to predict states directly, as the structure illustrated in Fig. 4.

Considering that the modeling accuracy of the active model in a specific application scene depends heavily on the adjustment of measurement noise, the covariance of the process noise is intentionally set to a small value to clearly evaluate the sensitivity of the different methods to measurement noise. Thus, the covariance of the process noise

TABLE II

RESULTS OF QUANTIFIED EVALUATION INDICES UNDER UNTRAINED DIFFERENT COVARIANCE OF GAUSSIAN NOISE

Covariance of measurement noise	Indices	C-H	AM	LSTM	**RAM**	Improvement[%]
$\boldsymbol{R_k} = 1$	MAE[deg]	4.4908	2.2051	1.9702	**0.4431**	90.1/ 79.9/ 77.5
	$RMSE$[deg]	5.4441	2.8922	2.4905	**0.6354**	88.3/ 78.8/ 74.4
	MD[deg]	2.5763	2.2200	1.9698	**0.4445**	82.7/ 79.9/ 77.4
	Var[deg^2]	9.6637	8.3489	6.2010	**0.4030**	95.8/ 95.1/ 93.5
$\boldsymbol{R_k} = 5$	MAE[deg]	6.6851	4.3098	2.3332	**0.7961**	88.1/ 81.5/ 65.9
	$RMSE$[deg]	8.7325	4.9759	2.8856	**1.0444**	88.0/ 79.0/ 63.8
	MD[deg]	4.8376	4.3393	2.3970	**0.7807**	84.6/ 82.0/ 67.4
	Var[deg^2]	12.2948	10.7782	6.3925	**1.0110**	91.8/ 90.6/ 84.2

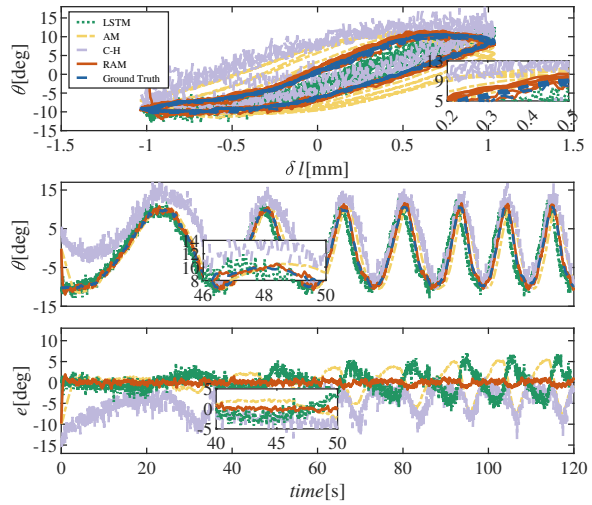

Fig. 6. The results in **Experiment 2:** $\boldsymbol{R_k} = 1$ (The ground truth values-blue dashed line; proposed scheme-orange solid line; the C-H model-purple dashed line; active model-yellow dash-dot line; LSTM-based model-green dotted line)

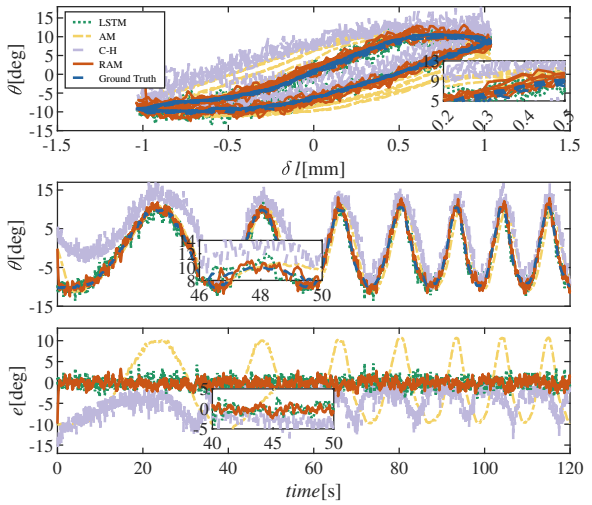

Fig. 7. The results in **Experiment 2:** $\boldsymbol{R_k} = 5$ (The ground truth values-blue dashed line; proposed scheme-orange solid line; the C-H model-purple dashed line; active model-yellow dash-dot line; LSTM-based model-green dotted line)

is set as $\boldsymbol{Q_k} = \mathbf{diag}\{0.001, 0.001\}$, while the covariance of the measurement noise in both training and testing are chosen as $\boldsymbol{R_k} = 0.1$ and $\boldsymbol{R_k} = 1$, respectively.

The results of **Experiment 1** for the test sets are depicted in Fig. 5. It is important to mention that an intentionally inaccurate initial estimation is provided at the onset of the modeling process. It can be seen whether Direct RL modeling or RAM can both quickly track the actual trajectory not affected by the inaccurate initial value. Moreover, it is noteworthy that Direct RL modeling exhibits a higher sensitivity to noise and larger modeling errors with sharp fluctuations compared to RAM. This can be attributed to the absence of a complete active modeling approach, which results in a loss of filtering capability. Thus, the intuitive superiority of the proposed scheme over Direct RL modeling can be demonstrated.

Experiment 2 *(generalization under different Gaussian noise):* To evaluate the generalization performance of the proposed scheme across varying covariance of Gaussian noise which is not introduced in training, experiments are conducted with the same training settings as the previous subsection, i.e., $\boldsymbol{Q_k} = \mathbf{diag}\{0.001, 0.001\}$ and $\boldsymbol{R_k} = 0.1$. Subsequently, generalization experiments are performed with two different cases of Gaussian noise, namely $\boldsymbol{R_k} = 1$ and $\boldsymbol{R_k} = 5$, to compare the modeling accuracy and modeling fluctuation of the proposed scheme with the C-H model, AM, and LSTM modeling scheme.

The results of **Experiment 2** are illustrated in Fig. 6 and Fig. 7. The proposed method consistently achieves rapid and satisfactory modeling accuracy with minor fluctuations. In contrast, for both $\boldsymbol{R_k} = 1$ and $\boldsymbol{R_k} = 5$, the C-H model, AM, and the LSTM modeling scheme exhibit significant modeling errors and noticeable oscillations. For the C-H model and AM, it required approximately 5 seconds to mitigate the interference of the inaccurate initial value in order to track the ground truth value. To objectively and quantitatively assess the modeling accuracy and fluctuation of the proposed and compared methods, four specific performance indices have been formulated as follows:

$$MAE = \frac{1}{N}\sum_{k=1}^{N}|e_k|,\ RMSE = \sqrt{\frac{1}{N}\sum_{k=1}^{N}e_k^2}$$
$$MD = \frac{1}{N}\sum_{k=1}^{N}\left|e_k^2 - \bar{e}_k^2\right|,\ Var = \frac{1}{N}\sum_{k=1}^{N}(e_k - \bar{e}_k)^2 \tag{13}$$

where e_k represents the estimation error of the proposed and comparative modeling schemes. MAE represents the mean absolute error of modeling accuracy, $RMSE$ is the root mean square error of modeling accuracy, MD represents the mean difference of modeling error, and Var is the variance of modeling error. MAE and $RMSE$ are introduced to

TABLE III

RESULTS OF QUANTIFIED PERFORMANCE INDICES UNDER UNTRAINED DIFFERENT EXPECTATIONS OF POISSON NOISE

Expectation of measurement noise	Performance indices	C-H	AM	LSTM	**RAM**	Improvement[%]
$\lambda = 0.1$	MAE [deg]	2.6982	0.6422	1.8049	**0.2768**	89.7/ 56.9/ 84.7
	$RMSE$ [deg]	3.7738	1.0049	2.2945	**0.4720**	87.5/ 53.0/ 79.4
	MD [deg]	2.3647	0.6612	1.7987	**0.2796**	88.2/ 57.7/ 84.5
	Var [deg^2]	7.7231	0.9913	1.8200	**0.2122**	97.3/ 78.6/ 88.3
$\lambda = 1$	MAE [deg]	3.5428	1.8271	2.1294	**1.0388**	70.7/ 43.1/ 51.2
	$RMSE$ [deg]	4.5736	2.5454	2.6640	**1.2125**	73.5/ 52.4/ 54.5
	MD [deg]	2.4925	1.7836	1.9693	**0.4549**	81.7/ 74.5/ 76.9
	Var [deg^2]	8.7495	5.2481	2.0690	**0.3970**	95.5/ 92.4/ 80.8
$\lambda = 2$	MAE [deg]	4.4908	2.7163	2.7476	**2.0156**	55.1/ 25.8/ 26.6
	$RMSE$ [deg]	5.4441	3.5148	3.3603	**2.1334**	60.8/ 39.3/ 36.5
	MD [deg]	2.5763	2.2095	2.1578	**0.5354**	79.2/ 75.8/ 75.2
	Var [deg^2]	9.6637	8.0237	2.4385	**0.4887**	94.9/ 93.9/ 80.0

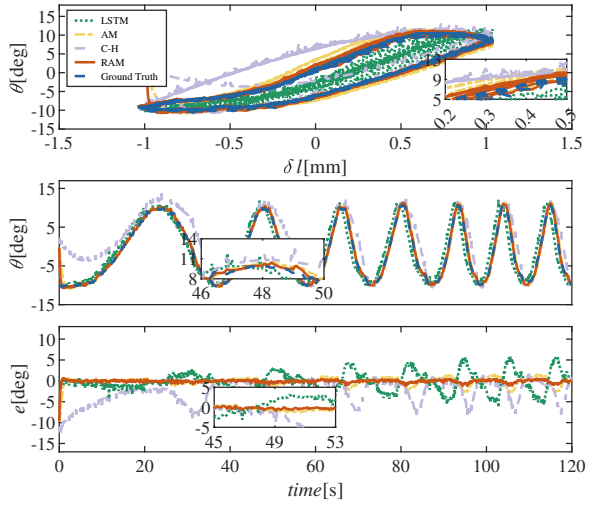

Fig. 8. The results in **Experiment 3:** $\lambda = 0.1$ (The ground truth values-blue dashed line; proposed scheme-orange solid line; the C-H model-purple dashed line; active model-yellow dash-dot line; LSTM-based model-green dotted line).

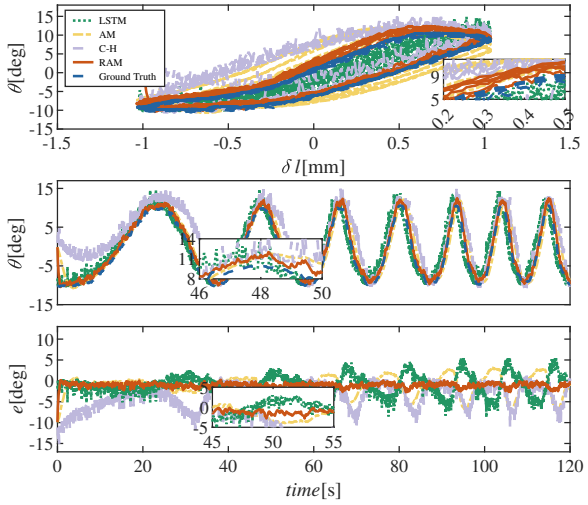

Fig. 9. The results in **Experiment 3:** $\lambda = 1$ (The ground truth values-blue dashed line; proposed scheme-orange solid line; the C-H model-purple dashed line; active model-yellow dash-dot line; LSTM-based model-green dotted line).

evaluate the modeling accuracy, and MD and Var are designed to describe the modeling fluctuation.

The quantified results of various modeling schemes are presented in Table II. It is evident that modeling errors and fluctuations increase when covariance $\boldsymbol{R_k}$ is larger. The proposed scheme consistently demonstrates outstanding performance across all performance indices and scenarios. It is noteworthy that RAM still exhibits an improvement of at least 78.8% in every index even compared with the learning-based LSTM modeling scheme. Specifically, compared to the C-H model, AM, and LSTM modeling scheme, when $\boldsymbol{R_k} = 1$, the proposed scheme exhibits a minimum reduction of 77.5% in MAE, 74.4% in $RMSE$, 77.4% in MD and 93.5% in Var.And when $\boldsymbol{R_k} = 5$, the proposed scheme presents a minimum reduction of 65.9% in MAE, 63.8% in $RMSE$, 67.4% in MD and 84.2% in Var. In summary, the outstanding modeling accuracy and satisfactory ability to mitigate the Gaussian noise of RAM have been validated by the experimental results.

Experiment 3 *(generalization under various non-Gaussian noise):* To verify the generalization capability of the proposed scheme suffering different non-Gaussian noise which is not introduced in training, the Poisson noise of the following distribution form is chosen as the measurement noise in this experiment:

$$P(x = i) = \frac{e^{-\lambda}\lambda^i}{i!} \tag{14}$$

where λ is the expectation. In this experiment, the expectations of the Poisson noise are set as $\lambda = 0.1$, $\lambda = 1$, and $\lambda = 2$, respectively. The experimental results are depicted in Figs. 8, 9, and 10. It can be seen that modeling errors and fluctuations increase when expectation λ is larger. In contrast, under various non-Gaussian noise, the proposed method consistently presents much more superior modeling accuracy and stability than comparative methods. The LSTM modeling scheme, despite being based on learning, exhibits unsatisfactory performance, occasionally even worse than AM in specific indices. Similarly, four performance indices defined as same as in **Experiment 2** are employed in this group. The quantified results are shown in Table III, indicating that when $\lambda = 0.1$, the proposed scheme

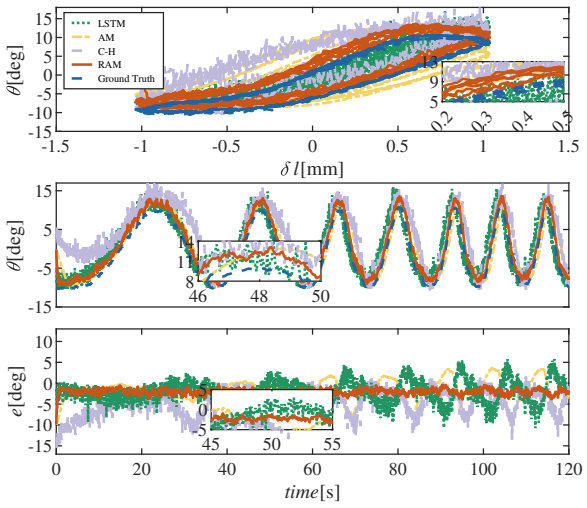

Fig. 10. The results in **Experiment 3:** $\lambda = 2$ (The ground truth values-blue dashed line; proposed scheme-orange solid line; the C-H model-purple dashed line; active model-yellow dash-dot line; LSTM-based model-green dotted line)

achieves a minimum reduction of 56.9% in MAE, 53.0% in $RMSE$, 58.7% in MD, and 78.6% in Var than comparative schemes, respectively. In particular, even under higher levels of Poisson noise expectation, these indices also show an improvement of more than 25% than others. In contrast, the LSTM modeling scheme consistently underperforms AM in most indices under non-Gaussian noise, indicating the limitations of its generalization capability. In contrast, the proposed scheme exhibits remarkable performance improvements of at least 25% over AM in all indices. In summary, the experimental result indicated that the proposed scheme has satisfactory modeling accuracy and remarkable capabilities to overcome non-Gaussian noise.

V. Conclusions

With respect to the concept of active estimation, a novel reinforcement active modeling scheme is introduced for flexible endoscopic robots in this paper. In particular, to overcome the limit of the Gaussian distribution assumptions and the dependency on the initial estimation accuracy of the reference model, this approach takes advantage of reinforcement learning and proposes a reinforcement active model to improve the modeling accuracy and the generalization ability of the model. Finally, a series of comparative experiments are conducted to validate the effectiveness of the proposed methodology. In future work, the robustness of the proposed active model will be further studied and utilized in real-time planning for flexible ureteroscopy robots.

References

[1] T. Yang, Y. Yang, P. Wang, Y. Cao, Z. Yang, and H. Liu, "A lumen-adapted navigation scheme with spatial awareness from monocular vision for autonomous robotic endoscopy," *Robotics and Autonomous Systems*, vol. 165, p. 104444, 2023.

[2] H. Gao, R. Hao, X. Yang, C. Li, Z. Zhang, X. Zuo, Y. Li, and H. Ren, "Modeling and compensation of stiffness-dependent hysteresis for stiffness-tunable tendon-sheath mechanism in flexible endoscopic robots," *IEEE Transactions on Industrial Electronics*, vol. 71, no. 8, pp. 9328–9338, 2024.

[3] Y. Zou, B. Guan, J. Zhao, S. Wang, X. Sun, and J. Li, "Robotic-assisted automatic orientation and insertion for bronchoscopy based on image guidance," *IEEE Transactions on Medical Robotics and Bionics*, vol. 4, no. 3, pp. 588–598, 2022.

[4] Y. Huang, W. Li, X. Zhang, J. Li, Y. Li, Y. Sun, P. W. Y. Chiu, and Z. Li, "4-DOF visual servoing of a robotic flexible endoscope with a predefined-time convergent and noise-immune adaptive neural network," *IEEE/ASME Transactions on Mechatronics*, vol. 29, no. 1, pp. 576–587, 2024.

[5] T. Do, T. Tjahjowidodo, M. Lau, and S. Phee, "Real-time enhancement of tracking performances for cable-conduit mechanisms-driven flexible robots," *Robotics and Computer-Integrated Manufacturing*, vol. 37, pp. 197–207, 2016.

[6] D.-H. Lee, Y.-H. Kim, J. Collins, A. Kapoor, D.-S. Kwon, and T. Mansi, "Non-linear hysteresis compensation of a tendon-sheath-driven robotic manipulator using motor current," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 1224–1231, 2021.

[7] X. Zhang, X. Chen, G. Zhu, and C.-Y. Su, "Output feedback adaptive motion control and its experimental verification for time-delay nonlinear systems with asymmetric hysteresis," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 8, pp. 6824–6834, 2020.

[8] B. Zhou, Y. Peng, and J. Han, "UKF based estimation and tracking control of nonholonomic mobile robots with slipping," in *2007 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pp. 2058–2063, IEEE, 2007.

[9] D. Song, J. Qi, J. Han, and G. Liu, "Active model based predictive control for unmanned helicopter in full flight envelope," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 616–621, 2010.

[10] F. Gu, Y. He, J. Han, and Y. Wang, "ESMF based multiple UAVs active cooperative observation method in relative velocity coordinates," in *48h IEEE Conference on Decision and Control (CDC)*, pp. 3008–3013, 2009.

[11] J. Xiong, D. Li, Y. He, F. Gu, and J. Han, "Active quasi-LPV modeling and identification for a water-jet propulsion USV: An experimental study," *IFAC-PapersOnLine*, vol. 48, no. 28, pp. 1359–1364, 2015.

[12] D. Zhang, X. Zhao, and J. Han, "Active model-based control for pneumatic artificial muscle," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 2, pp. 1686–1695, 2017.

[13] D. Zhang, X. Zhao, J. Han, X. Li, and B. Zhang, "Active modeling and control for shape memory alloy actuators," *IEEE Access*, vol. 7, pp. 162549–162558, 2019.

[14] X. Wang, D. Bie, J. Han, and Y. Fang, "Active modeling and compensation for the hysteresis of a robotic flexible ureteroscopy," *IEEE Access*, vol. 8, pp. 100620–100630, 2020.

[15] Y. Qin, H. Zhang, X. Wang, and J. Han, "Active model-based hysteresis compensation and tracking control of pneumatic artificial muscle," *Sensors*, vol. 22, no. 1, p. 364, 2022.

[16] X. Wang, N. Yu, D. Bie, J. Han, and Y. Fang, "A novel ESMF-based observer and control scheme for a type of tendon-sheath hysteresis system," *Automatica*, vol. 131, p. 109800, 2021.

[17] Y. Qin, H. Zhang, X. Wang, N. Sun, and J. Han, "Adaptive set-membership filter based discrete sliding mode control for pneumatic artificial muscle systems with hardware experiments," *IEEE Transactions on Automation Science and Engineering*, pp. 1–13, 2023.

[18] J. Chen, Y. Dang, and J. Han, "Offset-free model predictive control of a soft manipulator using the koopman operator," *Mechatronics*, vol. 86, p. 102871, 2022.

[19] L. Zheng, Z. Liu, Y. Wang, C. L. P. Chen, Y. Zhang, and Z. Wu, "Reinforcement learning-based adaptive optimal control for nonlinear systems with asymmetric hysteresis," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–10, 2023.

[20] D. Kim, H. Kim, and S. Jin, "Recurrent neural network with preisach model for configuration-specific hysteresis modeling of tendon-sheath mechanism," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 2763–2770, 2022.

[21] B. D. Coleman and M. L. Hodgdon, "A constitutive relation for rate-independent hysteresis in ferromagnetically soft materials," *International Journal of Engineering Science*, vol. 24, no. 6, pp. 897–919, 1986.

[22] D. Song, J. Han, and G. Liu, "Active model-based predictive control and experimental investigation on unmanned helicopters in full flight envelope," *IEEE Transactions on Control Systems Technology*, vol. 21, no. 4, pp. 1502–1509, 2012.