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

Human-Robot Interaction Control for Multi-Mode Exosuit with Reinforcement Learning

Kaizhen Huang, Jiajun Xu, *Member*, *IEEE*, Tianyi Zhang, Mengcheng Zhao,
Aihong Ji, Guoli Song, Youfu Li, *Fellow IEEE*

***Abstract*— Soft exoskeleton robots have promising potential in walking assistance with comfortable wearing experience. In this study, an exosuit equipped with a twisted string actuator (TSA) is developed to provide powerful driving force and diverse operating modes for hemiplegic patients in daily life. It is challenging to establish the human-robot coupling dynamic model due to the soft structure of the exosuit and tight coupling, precise control and effective assistance are difficult to guaranteed in current exosuits. Considering the impedance characteristics of human-robot interaction, an adaptive impedance control method based on reinforcement learning (RL) is proposed, where human motion intention is utilized to optimize impedance parameters and adjust the robot's operating mode. A nonlinear disturbance observer is proposed to compensate for the effects of model estimation errors, joint friction, and external disturbances. Experimental verification demonstrates the effectiveness and superiority of the robotic system.**

I. INTRODUCTION

Walking impairment is a common consequence of lower limb hemiparesis, affecting individuals who have experienced a stroke or neurological disorders. In recent years, the development of wearable robotic devices serves as a promising solution for assisting walking in patients with lower limb paralysis. [1-2]. These wearable devices hold the potential to augment residual motor function and improve gait biomechanics. The concept of "soft exosuits" was initially proposed by Harvard University, utilizing soft and flexible materials that are lightweight, enabling wearers to move more naturally [3-4]. Soft exosuits provide external assistance and support to impaired muscles, thereby enhancing the walking ability and quality of life for individuals with hemiparesis.

TSA is a novel actuator that utilizes Dynnema line as its driving element, which can adapt to various task requirements due to the robust output torque, high gear ratio, and adjustable stiffness. Moreover, compared with cable-driven systems, TSA offer a higher power-to-weight ratio, variable stiffness, and transmission ratios. Therefore, it is commonly used as a driving component in soft exosuits to assist wearers in performing movements. By establishing the kinematic model of the TSA, its stiffness, output torque, and gear ratio can be precisely controlled. However, throughout a complete motion cycle, the hysteresis phenomenon arising from the nonlinear characteristics of twisted strings (different pulling and releasing strokes during torsion) results in string recoil and tremors [5-6]. In this paper, we design the structure of TSA and evaluate its kinematic model to complement the effects of the physical properties of twisted strings.

Physical human-robot interaction control is crucial in rehabilitation robotics, especially for multi-mode applications [7-8]. Accurate establishment of human-robot coupling models is one of the necessary conditions for achieving precise control of exosuits. Due to the influence of material friction, tight coupling, and external disturbances, there are many parameters in the model that cannot be determined accurately. Some studies have considered the impedance characteristics during human-robot coupling, utilizing electromyography (EMG) to discern user intent, and proposing assist-as-needed strategies [9]. On the other hand, optimal parameters of the impedance model are obtained using integral RL to adjust the dynamics model and tracking performance [10]. Similarly, a model-based framework is employed to learn the assistance strategies of wearable exoskeletons, considering user voluntary efforts when designing the reward function [11-12].

To address the aforementioned issues, this paper proposes a soft exosuit driven by TSAs to provide walking assistance for individuals with hemiparesis and functions differently in multiple mode. An improved RL algorithm is introduced to optimize impedance parameters based on the muscle activity of patients, considering human-robot impedance characteristics. Controller design takes external disturbances into account and compensates for them using nonlinear observers. The specific contributions of this paper are as follows:

This work was supported in part by the National Natural Science Foundation of China under Grant 52205018, in part by the Natural Science Foundation of Jiangsu Province under Grant BK20220894, in part by the State Key Laboratory of Robotics and Systems (HIT) under Grant SKLRS-2023-KF-25, in part by State key Laboratory of Robotics under Grant 2023-O16, and in part by the China Postdoctoral Science Foundation under Grant 2024M754124. (*Corresponding author*: Jiajun Xu.)

Kaizhen Huang, Jiajun Xu, Tianyi Zhang, Mengcheng Zhao and Aihong ji are with the college of Mechanical and Electrical Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China (e-mail: huangkaizhen@nuaa.edu.cn; xujiajun@nuaa.edu.cn; zhangtianyi096@nuaa.edu.cn; zhaomengcheng@nuaa.edu.cn; meeahji@nuaa.edu.cn).

Guoli Song is with the State Key Laboratory of Robotics, Shenyang Institute of Automation, Shenyang 110169, China (e-mail: songgl@sia.cn).

Youfu Li is with the Department of Mechanical Engineering, City University of Hong Kong, Hong Kong, SAR, China (e-mail: meyfli@cityu.edu.hk).

(1) A soft exosuit with TSAs is designed and implemented to walking assistance for patient with hemiparesis in the daily life; and the kinematic model of the TSAs is derived.

(2) An optimization of impedance parameters is proposed using the enhanced TD3 algorithm, and multiple working modes are designed based on reward function.

(3) An adaptive impedance controller is designed to track the motion trajectory and compensate for external disturbances through nonlinear disturbance observe.

The remainder of this paper is organized as follows: the proposed exosuit with TSAs and torque estimation is described in Section II. The optimization of impedance parameter based on enhanced TD3 algorithm is explained in Section III. The adaptive control with nonlinear disturbance observer is introduced in Section IV. The experiments are conducted in Section V. Finally, the article is conducted in Section VI.

II. System Description

A Exosuit Overview

The presented soft exosuit in Fig. 1 is developed to provide knee joint assistance for hemiparetic patients with impaired lower limbs. Two sets of structurally identical TSAs are parallelly fixed on the wearer's backplate, transmitting torque to the wearer' s distal knee joint through Bowden cables. Typically, the rotation of the motor causes the twisted string to interlace and contract, tightening the Bowden cable, thereby pulling on the anchor point on the calf and inducing rotation at the knee joint. The Bowden cable, consisting of steel wire and a protective casing, is fastened to the other side of the connecting loop, passing through a guiding slot on the thigh and securing to an anchor point on the shin. In general, two TSAs collaborate to assist the flexion and extension of the lower limb knee joint. The design of the TSAs has been previously addressed in prior research [13]. This paper updates the structural design and kinematic model of the TSA.

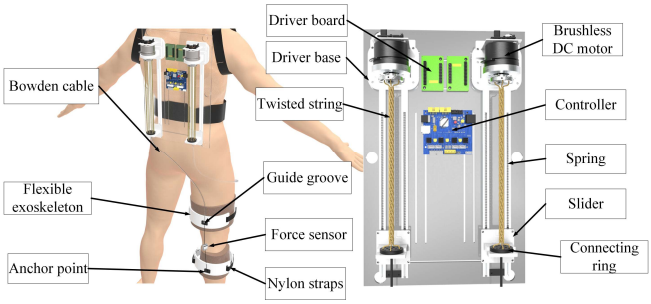

Fig 1. Overview of the Soft Exosuit.

B Kinematic Model

As shown in Fig.2(a), the twisted string winds and tightens under the action of the rotating motor, pulling the connecting ring to contract along the groove. In the previous research, we establish the constitutive relationship equation of the TSAs, as represented below:

$$F_h = nF_i = \frac{nK_h p}{L_0}\left(1 - \frac{L_0}{L}\right) \tag{1}$$

$$\tau_l = F_h \frac{q_m r_m^2}{p} = \frac{nK_h q_m r_m^2}{L_0}\left(1 - \frac{L_0}{L}\right) \tag{2}$$

where n is the strands of the twisted string, K_h is the strand stiffness of each twisted string, L_0 is the unloaded length of the strand, r_m is the string radius, p is the length of the twisted string transmission system, F_h is traction force output of twisted string, and $L = \sqrt{p^2 + q_m^2 r_m^2}$ is strand length.

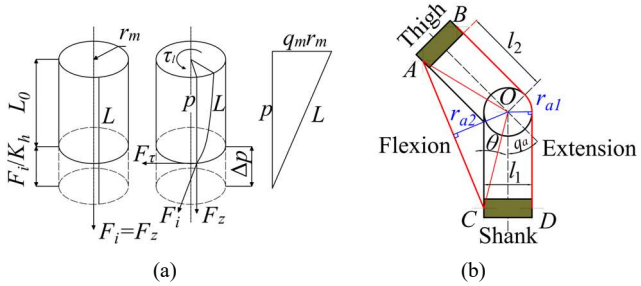

Fig 2. The schematic of kinematic model. (a) Geometry of the string during the twisted; (b) The schematics of the tendons routing on the knee.

Considering the friction losses resulting from the nonlinear characteristics of the Bowden cable during operation, along with the pre-tension force of the spring, the Bowden cable output tension vector F_a acting on the anchor point of the lower leg can be expressed as:

$$F_a = F_h - F_k - F_f \tag{3}$$

$$F_k = \sigma \Delta X = \sigma(L - p) \tag{4}$$

$$F_f = C\mathrm{sgn}(\dot{q}_m)\dot{q}_m \tag{5}$$

where F_f represents the friction losses of the Bowden string, respectively; F_k represent the pre-tension force generated by the spring; σ is the friction factor between the strands; ΔX represent the linear displacement generated by the twisted string. Therefore, the relationship between the Bowden cable output traction force F_a and the knee joint torque vector τ_a can be expressed as:

$$\tau_a = R_a F_a \tag{6}$$

where $R_a = \mathrm{diag}[r_{a1}, r_{a2}] \in \mathbb{R}^{2\times 2}$ represent the vertical distance between the knee joint centers and Bowden cable, the vertical distance from the knee joint to the Bowden cable on different sides can be expressed as:

$$R_a = \begin{cases} \lambda \sin\left(\frac{q_a}{2} + \theta\right), & \text{Flexion side} \\ r_e, & \text{Extension side} \end{cases} \tag{7}$$

where $\theta = arctan\frac{l_1}{2l_2}$, $\lambda = \sqrt{{l_2}^2 + \frac{1}{4}{l_1}^2}$, l_1 is the widths of the thigh and calf, l_2 is the distance between the anchor points and knee joint, q_a represents the knee joint angular position, refer to Fig. 2(b) for specific details. Thus, from (1)-(7), we can obtain:

$$\tau_l = \frac{q_m R_m^2}{p}\left(Csign(\dot{q}_m)\dot{q}_m + \sigma\left(\sqrt{p^2 + q_m^2 r_m^2} - p\right) + R_a^{-1}\tau_a\right) \tag{8}$$

From (8), we have accounted for the pre-tension force of the spring, frictional losses between the Bowden cable and the sleeve, establishing a kinematic model relating the rated torque of the motor to the output torque of the soft exosuit. The actuator stiffness of the TSA can be characterized as $\partial\tau_a/\partial q_m$, and it can be modulated by altering the structure

of the TSA and the output configuration of the motor to accommodate the assistance requirements of various wearers.

C Torque Estimation

In designing the controller for soft exosuits, the motion trajectory and stiffness of the exosuit should be adjusted based on the subject's voluntary torque and stiffness to enhance the assistance effect. Specifically, the output torque and stiffness of the TSAs should adapt to the muscle force of the subject's IL. In this section, the voluntary torque of the subjects is assessed based on surface EMG signals. The collected EMG signals are filtered and normalized, and the post-processed EMG signal is transformed to the muscle activation by:

$$a_i(t) = \frac{e^{Au_i(t)-1}}{e^{A}-1} \tag{9}$$

where $u_i(t)$ is the amplitude of the pre-processed EMG signal for the ith muscle, $a_i(t)$ is the activation of the ith muscle, and $A \in (-3,0)$ is the nonlinear shape factor.

Then an effectively validated musculoskeletal model is used to calculate the joint torque and stiffness of the human body [14]:

$$\tau_e(t) = \left|\sum_{i=1}^{n} \tau_i(t)\right|_{agonist} - \left|\sum_{j=1}^{n} \tau_j(t)\right|_{antagonist} \tag{10}$$

where $\tau_i(t) = F_i^m r_i^m$ and $\tau_j(t) = F_j^m r_j^m$ represent the joint torque exerted by the agonistic muscle (rectus femoris; vastus medialis; vastus lateralis) and antagonistic muscles (biceps femoris; semimembranosus; semitendinosus), respectively; F_i^m and F_j^m are the muscle force of the corresponding musculotendinous unit; r_i^m and r_j^m are the moment arms of the corresponding muscle-tendon unit. Subsequently, the human joint stiffness can be expressed as:

$$K_e(t) = \alpha_j\left(\left|\sum_{i=1}^{n} \tau_i(t)\right|_{ag} + \left|\sum_{i=1}^{n} \tau_i(t)\right|_{\text{ant}}\right) + \beta_j \tag{11}$$

where $\alpha_j\ (rad^{-1})$ and $\beta_j\ (Nm/rad)$ are to be identified constants. To calibrate α_j and β_j, conduct the maximum voluntary contraction experiment of muscles to compare the measured joint stiffness and the estimated joint stiffness among different subjects, aiming to minimize estimation errors [9]. Note that the estimated joint stiffness on the impaired side here is not used to replace parameters in the impedance model but serves as an evaluation metric for RL to optimize impedance parameters, which will be described in the next section.

III. Reinforcement Learning

A Impedance Model

The soft exosuit constitutes an intricately coupled intelligent system, where one of its assessment criteria lies in providing the knee joint with an optimally tailored assistive effect during subject motion. In designing the controller for the soft exosuit, adjustments to its motion trajectory and stiffness are made based on the subject's voluntary torque and rigidity, ensuring safety while enhancing assistance efficacy. Specifically, the output torque and stiffness of the TSAs should accommodate the muscular forces of the subject's IL. In this section, we explore the impedance characteristics between the knee joint of the IL and the soft exosuit. It's crucial to note that unforeseen changes in inertia may result in model instability. Hence, the impedance model omitting the inertia matrix can be represented as follows:

$$\overline{B}(\dot{q}_d - \dot{q}_r) + \overline{K}(q_d - q_r) = \tau_{IL} \tag{13}$$

where $\overline{B}$ and $\overline{K}$ are desired damping and stiffness, respectively; the desired trajectories and reference trajectories of the knee joint is denoted as q_d and q_r, respectively. To obtain the reference trajectories, the unknown impedance parameter $\overline{B}$ and $\overline{K}$ need to be estimated.

B Impedance Learning

The traditional method of adjusting impedance parameters typically requires extensive experimentation and expert experience. In contrast, RL algorithms continuously adjust and optimize parameters through a trial-and-error and reward mechanism. This learning mechanism makes impedance parameters dependent on the configuration of the reward function, such as EMG signals and joint angles, rather than relying on fixed parameters. Deep deterministic policy gradient algorithm addresses RL probiems in continuous action spaces, but it may encounter local optima and estimation issues in complex tasks. Inaccurate value estimation of the reward function can lead to inefficient sample training and even serious safety concerns. In this paper, the TD3 algorithm is employed to update training parameters and enhance the algorithm's performance, the network architecture of TD3 is shown in Fig.3.

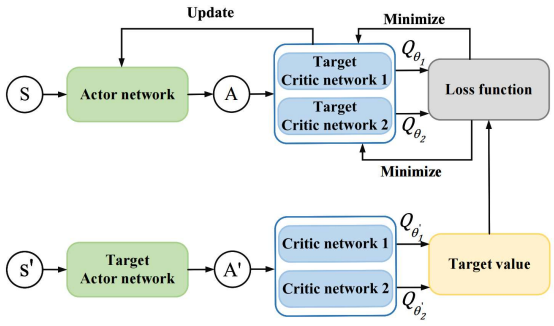

Fig 3. The network architecture of TD3.

As described in the Markov decision model, the robot's future state depends on its current state, the subject's current state, and the actions taken by the robot. To achieve safe and stable tracking performance, the state of the exosuit primarily includes joint positions and velocities. Note that we preset motion range and joint torque limits as learning hyperparameters to ensure user safety. Muscle activation of patients is one of the indicators of the assistive effect. Generally, due to limited mobility, muscle activation in the IL is significantly weaker than in the HL, reflected in joint torque and stiffness. Considering increased joint stiffness in spasticity, while torque remains constant, joint stiffness is incorporated into the human state to ensure patient safety [15]. Therefore, the RL agent utilizes multiple indicators to evaluate system status, represented as:

$$s = [q, \dot{q}, K_{HL}, K_{IL}] \tag{14}$$

Furthermore, the exosuit's action is the unknown impedance parameters of the human-robot coupling model, denoted as:

$$a = [\overline{K}, \overline{B}] \tag{15}$$

The objective of optimizing impedance parameters through RL aim to minimize tracking errors of the soft exosuit and reduce the user's effort in motion tasks. The reward function should be determined by the following two rules:

Rule 1: Maximizing the reduction of trajectory errors between the HL and IL. Ensuring the stable tracking of HL's motion trajectory by IL to guarantee safety and comfort.

Rule 2: Maximizing the enhancement of assistance provided to the IL. Elevating joint stiffness in patients actively engaged in gait training to prevent sudden exertion and subsequent spasms.

According to these roles, the multi-objective reward function is formulated as:

$$r = -(q - q_d)^T \Lambda_1 (q - q_d) - (\dot{q} - \dot{q}_d)^T \Lambda_2 (\dot{q} - \dot{q}_d) + \Lambda_3 \frac{K_{IL}}{K_{HL}} \tag{16}$$

where Λ_1 and Λ_2 denote the weighting of the tracking error component in the reward function, and Λ_3 is a constant that balances the contribution of the joint stiffness, the weights of each parameter can be determined through the relative entropy inverse RL algorithm [16].

The pseudocode for learning is shown in Algorithm 1. when updating the Q-value using the Bellman operator, the updated Q-value, denoted as $[r(s,a) + Q(s',a')]$, is expected to be more accurate than the original Q-value. Therefore, comparing it with the updated Q-value allows us to determine whether the current state is overestimated or underestimated, it can be reppresneted as:

$$\zeta = Q(s,a) - [r(s,a) + Q(s',a')] \tag{17}$$

where ζ is an adaptive evaluation metric, $r(s,a)$ is the instantaneous reward under state s and action a. Then, regularizing ζ based on existing rewards, which enables adaptive adjustment across different scales of reward functions, as shown in:

$$\bar{\zeta} = \frac{\zeta}{\max(r_1, r_2, .., r_t)} \tag{18}$$

where t represents the current time step. From (24), it is evident that the reward function is closely related to the subject's joint stiffness. There are three modes in human-robot interaction: in the robot-active mode, the robot guides the subject to complete trajectory tasks, and the people just needs to follow the robot's motion without exerting additional muscle force; in the normal assistance mode, torque is collaboratively generated by both the human and the robot to accomplish the gait task; and in the human-active mode, the robot becomes compliant and passive, supplementing the subject's missing torque to maximize the patient's muscle strength. In our control objective, different value estimation results correspond to different modes of human-robot interaction. When $\bar{\zeta} > \zeta_1$, it signifies the robot-active mode, where the Q-function consistently selects actions that maximize value in the current state, i.e.: $Q^{\zeta}(s,a) = \max[Q_1(s,a), Q_2(s,a)]$. The robot exhibits higher trajectory tracking accuracy and provides greater assisting torque. When $\bar{\zeta} < \zeta_2$, it signfies the human-active mode, where the robot assists the human in completing the gait task to maximizing activation at the patient's muscle level, i.e.: $Q^{\zeta}(s,a) = \min[Q_1(s,a), Q_2(s,a)]$. When $\zeta_2 < \bar{\zeta} < \zeta_1$, it signfies normal assistance, the gait task is completely bu the robot and human, without distinction of dominance, the Q-value is calculated as:

$$Q^{\zeta}(s,a) = |\zeta| \max[Q_1(s,a), Q_2(s,a)] + (1 - |\zeta|) \min[Q_1(s,a), Q_2(s,a)] \tag{19}$$

Note that the parameters ζ_1 and ζ_2 determine the threshold for numerical estimation and can be adaptively adjusted according to the assistance requirements of different subjects.

Algorithm 1: Impedance parameters learning based on the ζ-TD3 methods.

Initialize: critic networks Q_{θ_1}, Q_{θ_2}, and normalized actor network π_{ϕ} with random parameters θ_1, θ_2, ϕ
Initialize: target networks $\theta_1' \leftarrow \theta_1$, $\theta_2' \leftarrow \theta_2$, $\phi' \leftarrow \phi$
Initialize: replay buffer $\boldsymbol{\mathcal{B}}$
For episode = 1 **to** P **do**
  Receive the initial state S_1 by measuring the exosuit trajectory positions, velocity, and EMG signals through the sensor
  Calculate the muscle activation A_{HLi}, A_{ILi}, and joint stiffness K_{HL}, K_{IL}
  For t = 1 **to** T **do**
    Select action $a' = \mu'(s'|\theta^{\mu'}) + \epsilon$, $\epsilon \sim N(0, \sigma)$, and observe reward r, and receive a new observation state s'
    Store transition tuple (s, a, r, s') in buffer $\boldsymbol{\mathcal{B}}$
    Sample minibatch of N transitions (s, a, r, s') from buffer $\boldsymbol{\mathcal{B}}$
    Set $y_i = r_i + \gamma Q^{\zeta}(s,a)$
    Update critic θ_i by minimizing the loss:
    $L(\theta_i) = N^{-1} \sum (Q_{i=1,2}(s,a) - y_t)^2$ (20)
    If t mod d **then**
      Update ϕ by deterministic policy gradient:
      $\nabla_{\phi} J(\phi) = N^{-1} \sum \nabla_a Q_{\theta 1}(s,a)|_{a=\pi_{\phi}(s)} \nabla_{\phi} \pi_{\phi}(s)$
      Update target networks:
      $\theta_i' = \sigma\theta_i + (1 - \sigma)\theta_i'$
      $\phi' = \sigma\phi + (1 - \sigma)\phi'$ (21)
    end if
  end for

IV. Control Method

A Dynamic Model

Consider the interaction torques between human and exosuit, the dynamic model of the human-robot coupling system can be expressed as:

$$M(q)\ddot{q} + C(q,\dot{q})\dot{q} + G(q) + D = \tau_{act} + \tau_{IL} \tag{22}$$

where $q \in \mathbb{R}^n$ is the position coordination of the robotic joint, $\dot{q}$ and $\ddot{q}$ represent the joint velocity and acceleration

respectively. $M(q) \in \mathbb{R}^{n\times n}$, $C(q,\dot{q}) \in \mathbb{R}^{n\times n}$, and $G(q) \in \mathbb{R}^n$ are the symmetric matrix, Centripetal and Coriolis matrix, and gravitational torque, respectively. D is the external disturbance. $\tau_{act} \in \mathbb{R}^n$ is the robotic joint torque generated by the TSAs, and $\tau_{IL} \in \mathbb{R}^n$ is the voluntary torque exerted by the IL.

Then, some important properties of the dynamic model are used to design controller and analyze the stability.

Property 1: The matrix $\dot{M}(q) - 2C(q,\dot{q})$ is skew symmetric.

Property 2: $M(q)$ in (22) is positive definite symmetric, with the following property:

$$m_1\|\xi\|^2 \le \xi^T M(q)\xi \le m_2\|\xi\|^2 \forall \xi \in \mathbb{R}^n \tag{23}$$

Assumption 1: The disturbance D is assumed to be continuous because it can be largely attributed to the material friction, estimation error, and gravity; D has finite energy, i.e.: $\|\dot{D}\| \le d_m$, where d_m is an unknown positive constant.

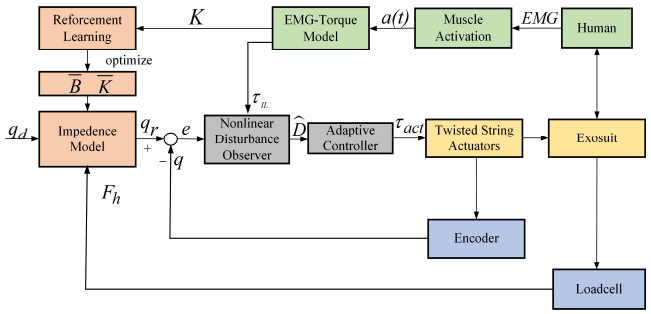

Fig 4. Control diagram for the dual-loop design.

B Controller Design

From (22), $M(q)$, $C(q,\dot{q})$, $G(q)$,and D are unknown parameters, it is challenging to ensure that the system's output can track the reference trajectory q_r in the presence of uncertainties and external disturbances. Therefore, inspired by [17], we address the issue of compensating for joint friction and external disturbances using a nonlinear observer. Additionally, considering the impedance characteristics between the knee joint and the soft exosuit during motion, we propose an improved RL algorithm, which optimizes the impedance model based on the subject's performance, iteratively seeking the optimal assistance parameters through learning.

We define the following error variable $e_1 \in \mathbb{R}^n$ and $e_2 \in \mathbb{R}^n$ as:

$$e_1 = q - q_r \tag{24}$$

$$e_2 = \dot{q} - \alpha_1 \tag{25}$$

where α_1 is virtual control to e_1. Therefore, the derivatives of e_1 and e_2 can be expressed as:

$$\dot{e}_1 = \dot{q} - \dot{q}_r = e_2 + \alpha_1 - \dot{q}_d \tag{26}$$

$$\dot{e}_2 = M(q)^{-1}[\tau_{act} + \tau_{IL} - C(q,\dot{q})\dot{q} - G(q) - D] - \dot{\alpha}_1 \tag{27}$$

We assume that D is continuous because it can largely be attributed to exogenous effects, D has finite energy and satisfies, i.e., $\|\dot{D}\| \le d_m$, and d_m is an unknown positive constant. Then, a nonlinear disturbance observe can be designed to estimate the disturbance D and the following auxiliary can be defined as:

$$e_3 = D - \phi(e_2) \tag{28}$$

where $\phi(e_2)$ is a nonlinear function vector $\phi(e_2) \in \mathbb{R}^n$, combine (27) with (28), we can obtain the derivative of e_3 as:

$$\dot{e}_3 = \dot{D} - K(e_2)M(q)^{-1}[\tau_{act} + \tau_{IL} - C(q,\dot{q})\dot{q} - G(q) - D] - K(e_2)\dot{\alpha}_1 \tag{29}$$

where $K = \partial\phi(e_2)/\partial e_2 \in \mathbb{R}^n$. To estimates the nonlinear disturbance in the system using an observe, the estimate $\dot{\hat{e}}_3$ is given by:

$$\dot{\hat{e}}_3 = K(e_2)M(q)^{-1}\left[\tau_{act} + \tau_{IL} - C(q,\dot{q})\dot{q} - G(q) - \hat{D}\right] - K(e_2)\dot{\alpha}_1 \tag{30}$$

where $\hat{e}_3$ is the estimate of the e_3. Then, we can have the estimate of the disturbance D and the error estimate $\tilde{e}_3$.

$$\hat{D} = \hat{e}_3 + \phi(e_2) \tag{31}$$

$$\tilde{e}_3 = e_3 - \hat{e}_3 = D - \hat{D} = \tilde{D} \tag{32}$$

Combining (29) and (30) and then differentiating (32),we obtain:

$$\dot{\tilde{D}} = \dot{e}_3 - \dot{\hat{e}}_3 = \dot{D} - K(e_2)M(q)^{-1}\tilde{D} \tag{33}$$

Then, the controller can be design as:

$$\tau_{act} = M(q)\dot{\alpha}_1 + C(q,\dot{q})\alpha_1 + G(q) + \hat{D} - e_1 - \varepsilon e_2 - \tau_{IL} \tag{34}$$

where ε is the positive gain. Considering the Lyapunov function cadidate:

$$V = \frac{1}{2}e_1^T e_1 + \frac{1}{2}e_2^T M(q)e_2 + \frac{1}{2}\tilde{D}^T\tilde{D} \tag{35}$$

Let $\alpha_1 = \dot{q}_r - \delta e_1$, according to ***property 1***, the time derivative of Lyapunov function cadidate is:

$$\begin{aligned}\dot{V} = &-e_1^T\delta e_1 + e_1^T e_2 \\ &+e_2^T \times [\tau_{act} + \tau_{IL} - C(q,\dot{q})\alpha_1 - G(q) - D \\ &+\frac{1}{2}\left(\dot{M}(q) - 2C(q,\dot{q})\right)e_2] + \tilde{D}^T\dot{\tilde{D}}\end{aligned} \tag{36}$$

Combining (34) with 36), we can obtain:

$$\begin{aligned}\dot{V} \le &-e_1^T\delta e_1 - e_2^T\left(\varepsilon - \frac{1}{2}I\right)e_2 \\ &-\tilde{D}^T(K(e_2)M(q)^{-1} - I)\tilde{D} + \frac{1}{2}d_m^2\end{aligned} \tag{37}$$

According to the Lyapunov stability criteria, when the controller parameters δ and ε are positive, $\varepsilon - \frac{1}{2}I > 0$, $K(e_2)M(q)^{-1} - I > 0$, the Lyapunov candidate function is positive definite, its time derivative is negative definite, and the system stability is proved.

V. Experiment

In this section, the proposed controller is experimentally validated for trajectory tracking performance, performance of the nonlinear disturbance observer, and convergence of reinforcement learning. Five hemiparetic patients participate

in the experiment after being informed of the detailed surgical procedure and potential risks, and they sign informed consent forms. No severe spasms occur during the experiment, and the experimental setup is depicted in Fig.4. An embedded control board (STM32H764) communicates with the actuator board (EPOS4 50) via the CAN bus to control two brushless DC motors (Maxon EC60 Flat). Additionally, a semi-physical real-time simulation system (CSPACE) serves as the host computer, interacting with the control board in real time through the CAN network. The functionality of this interface can be understood as utilizing dynamic link libraries to communicate with the control board (CANOpen protocol wired CAN bus) and monitor signals in real time during the execution of control strategies. Surface EMG acquisition system (Biometrics Ltd, WS250) and IMU (WitMotion) sensors are adhered to designated muscles on the subjects' thighs, as shown in Fig. 6. Muscle relaxation and alcohol cleaning are necessary before EMG signal acquisition. The raw EMG signals are sampled at 2 kHz and filtered to remove noise.

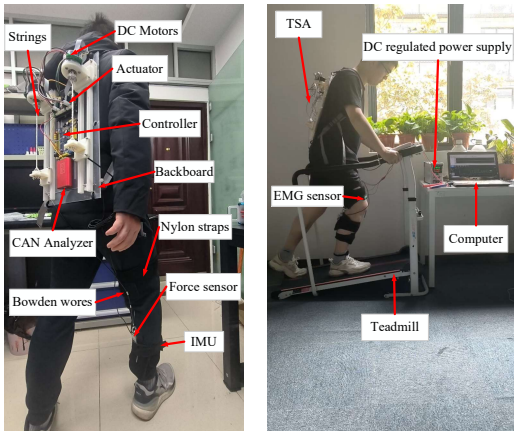

Fig 5. Robot prototype.

Fig 6. Collection of electromyography signals and joint angle signals.

During the experiment, the HL of the patients performs gait movements under the guidance of professional physiotherapists, serving as the desired trajectory for control design. Simultaneously, the IL of the patients completes corresponding gait tasks with the assistance of the soft exosuit driven by the adaptive impedance controller introduced in the paper. To reduce random errors in single measurements, each subject undergoes three experiments, and the averages are calculated to ensure accuracy. The desired trajectory (q_d), the reference trajectory of the impedance model (q_r), and the actual trajectory (q_a) of the IL with exosuit assistance, along with their corresponding trajectory velocities, are illustrated in Fig.7. Generally, due to external disturbances, model uncertainties, and individual differences, the output trajectory of the impedance model cannot perfectly track the desired trajectory. It can be seen that the tracking effect of these three trajectories are decent. In the initial stages of gait, inaccurate impedance parameters affect the performance of the controller. After interaction training between the human body and the exoskeleton, the proposed reinforcement learning algorithm can optimize impedance parameters based on the state information of the IL, thereby achieving optimal trajectory tracking effects. Furthermore, the proposed controller compensates for error disturbances caused by factors such as estimation errors, joint friction, and gravity using the nonlinear disturbance observer, as illustrated in Fig.8.

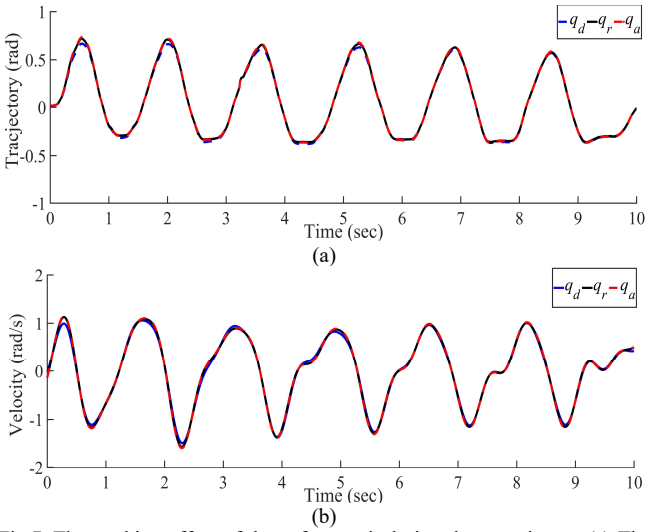

Fig 7. The tracking effect of the soft exosuit during the experiment. (a) The position of the desire trajectory, reference trajectory and actual trajectory; (b) The velocity of the desire trajectory, reference trajectory and actual trajectory.

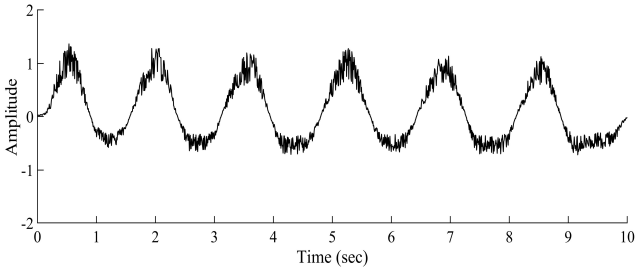

Fig 8. Designed nonlinear disturbance observers.

In addition to demonstrating excellent trajectory tracking performance, the proposed controller must also exhibit outstanding assistive effects. Fig. 9 depicts the EMG signals and joint torque of different muscles in the IL of the subjects, Fig.9(a) representing the case without wearing the soft exosuit, and Fig. 9(b) representing the case with wearing the soft exosuit. Muscle activation ensures active participation of the patient in the gait task, rather than passive reliance on the exosuit for movement completion. It is worth noting that the proposed soft exosuit demonstrates commendable assistance, with the reinforcement mechanism optimizing the parameters of the impedance model, thereby enhancing the exosuit's assisting capability. Even in cases where initial impedance model parameters differ among subjects, the proposed RL controller can explore the most optimal assistive strategies tailored to individual's specific gait tasks.

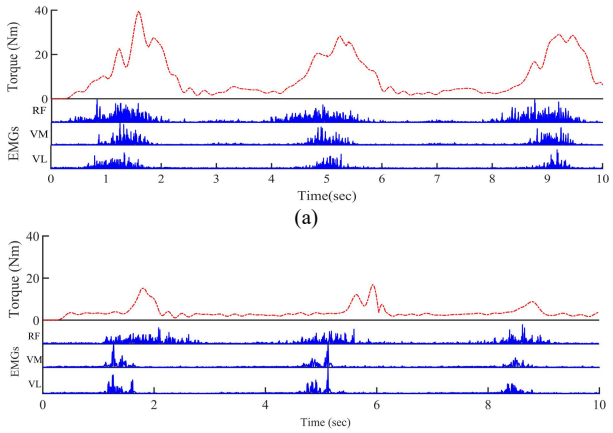

Fig 9. Comparison of normalized EMG signals of different thigh muscles and joint torque before and after wearing the soft exosuit. (a) EMG signals and joint torque without wearing the soft exosuit; (b) EMG signals and joint torque with wearing the soft exosuit.

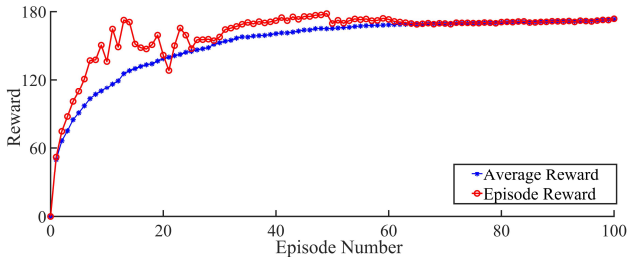

Fig 10. The episode and average reward versus the episode number during the RL process.

As mentioned above, to assist patients more effectively in actively completing walking tasks, the proposed RL algorithm underwent meticulous refinement for each participant. Therefore, assessing the learning efficiency of the proposed controller is imperative. During the training process, certain hyperparameters will be set, namely the learning rates $\sigma_1 = 0.0001$ and $\sigma_2 = 0.001$for the critic and actor networks respectively, the experience replay buffer $R = 10^6$, and the discount factor $\gamma = 0.995$. The rewards per episode and the average reward over episodes are illustrated in Fig. 10. Following the initiation of training, rewards rapidly escalate and converge around the 50th episode. Ultimately, the average reward value reaches 170, requiring approximately 1 hour to accomplish. By assisting multiple patients with the exosuit, thereby collecting more data, updating gradient policies, and optimizing algorithms, further reduction in training time can be achieved.

Compared to other state-of-art control strategies, the proposed RL controller aims to maximally assist the IL in tracking the HL and enhance its muscle activation. Additionally, the inclusion of the nonlinear disturbance observer in this framework allows for the elimination of uncertainties in robot dynamics. Overall, this approach demonstrates advantages in ensuring safety, assistance effectiveness, and optimization efficiency.

VI. CONCLUSION

To assist individuals with hemiplegia in walking, this paper proposes and designs a soft exosuit actuated by TSAs, whose drive stiffness and working modes can adaptively adjust to accommodate patients with various functional abilities. An adaptive impedance control method based on a dual-delay deep RL algorithm is introduced, offering safe and effective assistance under conditions of limited information on human-exosuit coupled dynamics. Experimental validation demonstrates the effectiveness and superiority of the proposed flexible exosuit system.

In future work, considering the impact of the environment on human-robot interaction systems during walking in everyday life, we will incorporate visual sensors to effectively capture environmental information surrounding the system. Furthermore, more intelligent control methods will be proposed to further enhance the assistive capability.

REFERENCES

[1] Kim. J *et al.*, "Reducing the energy cost of walking with low assistance levels through optimized hip flexion assistance from a soft exosuit." *Sci Rep*, vol. 12, pp. 11004-11017, 2022.

[2] X. Wang *et al.*, "Design and experimental verification of a hip exoskeleton based on human-machine dynamics for walking assistance," *IEEE T. Hum. -Mach. Syst.*, vol. 53, pp. 85-97, 2023.

[3] A. T. Asbeck *et al.*, "A biologically inspired soft exosuit for walking assistance," *Int. J. Robot. Res.*, vol. 34, no. 6, pp. 744-762, 2015.

[4] Y. Ding *et al.*, "Biomechanical and physiological evaluation of multi-joint assistance with soft exosuits," *IEEE Trans. Neural Syst. and Rehab. Eng.*, vol. 25, no. 2, pp. 119-130, 2017.

[5] G. Palli *et al.*, "Twisted string actuation with sliding surfaces," *2016 IEEE Int Conf. Intel. Rob Syst*, Daejeon, Korea, 2016, pp. 260-265.

[6] B. Suthar *et al.*, "Preliminary Study of Twisted String Actuation Through a Conduit Toward Soft and Wearable Actuation," *2018 IEEE Int Conf. Rob. and Autom.,* Brisbane, Australia, 2018, pp. 2260-2265.

[7] X, Li *et al.*, "Adaptive human–robot interaction control for robots driven by series elastic actuators." *IEEE Trans. Robot.*, vol.33, no.1 pp.169-182, 2017.

[8] X. Li *et al.*, "Multi-modal control scheme for rehabilitation robotic exoskeletons," *Int. J. Robot. Res.*, vol. 36, nos. 5–7, pp. 759–777, Jun. 2017.

[9] J. Xu *et al.*, "A Multi-Mode Rehabilitation Robot with Magnetorheological Actuators Based on Human Motion Intention Estimation," *IEEE Trans. Neural Syst. and Rehab. Eng.*, vol. 27, no. 10, pp. 2216-2228, 2019.

[10] H. Modares *et al.*, "Optimized assistive human-robot interaction using reinforcement learning," *IEEE Trans. Cybern.*, vol. 46, no. 3, pp. 655–667, Mar. 2016.

[11] M. Hamaya *et al.*, "Learning assistive strategies from a few user-robot interactions: Model-based reinforcement learning approach," in Proc. *IEEE Int. Conf. Robot. Autom.*, Stockholm, Sweden, May 16-21, 2016, pp. 3346–3351.

[12] J. Xu *et al.*, "A Multi-Channel Reinforcement Learning Framework for Robotic Mirror Therapy," *IEEE Rob. Autom. Lett.*, vol. 5, no. 4, pp. 5385-5392, 2020.

[13] K. Huang *et al.*, "A Variable stiffness exosuit for walking assistance," *J. Phys.: Conf. Ser.*, vol. 2402, pp. 012033, 2022.

[14] R. Osu *et al.*, "Short- and Long-Term Changes in Joint Co-Contraction Associated with Motor Learning as Revealed from Surface EMG," *J. Neurophysiology.*, vol 88, pp. 991-1004, 2002.

[15] J. M. Hidler *et al.*, "Strength and Coordination in the Paretic Leg of Individuals Following Acute Stroke," *IEEE Trans. Neural Syst. and Rehab. Eng.*, vol. 15, no. 4, pp. 526-534, 2007.

[16] A. Boularias *et al.*, "Relative entropy inverse reinforcement learning," in *Proc. Artif. Intell. Statist.*, 2011, pp.20-27.

[17] Z. Li *et al.*, "Nonlinear Disturbance Observer-Based Control Design for a Robotic Exoskeleton Incorporating Fuzzy Approximation," *IEEE Trans. Ind. Electron.*, vol. 62, no. 9, pp. 5763-5775, 2015.