processes

MDPI

Article

Intelligent Impedance Strategy for Force–Motion Control of Robotic Manipulators in Unknown Environments via Expert-Guided Deep Reinforcement Learning

Hui Shao [1,2,*], Weishi Hu [3], Li Yang [1,2], Wei Wang [4,5], Satoshi Suzuki [5] and Zhiwei Gao [6,*]

[1] College of Information Science and Engineering, Huaqiao University, Xiamen 361021, China
[2] Fujian Engineering Research Center of Motor Control and System Optimal Schedule, Xiamen 361021, China
[3] Laboratory and Equipment Management Department, Huaqiao University, Xiamen 361021, China
[4] Autonomous, Intelligent, and Swarm Control Research Unit, Fukushima Institute for Research, Education and Innovation (F-REI), Futaba County, Fukushima 979-1521, Japan; wang.wei.r4k@research.f-rei.go.jp
[5] Graduate School of Engineering, Chiba University, 1-33 Yayoi-cho, Inage-ku, Chiba-shi, Chiba 263-8522, Japan
[6] Faculty of Engineering and Environment, University of Northumbria at Newcastle, Newcastle upon Tyne NE1 8ST, UK
* Correspondence: shaohuihu11@163.com (H.S.); zhiwei.gao@northumbria.ac.uk (Z.G.)

Abstract

In robotic force–motion interaction tasks, ensuring stable and accurate force tracking in environments with unknown impedance and time-varying contact dynamics remains a key challenge. Addressing this, the study presents an intelligent impedance control (IIC) strategy that integrates model-based insights with deep reinforcement learning (DRL) to improve adaptability and robustness in complex manipulation scenarios. The control problem is formulated as a Markov Decision Process (MDP), and the Deep Deterministic Policy Gradient (DDPG) algorithm is employed to learn continuous impedance policies. To accelerate training and improve convergence stability, an expert-guided initialization strategy is introduced based on iterative error feedback, providing a weak-model-based demonstration to guide early exploration. To rigorously assess the impact of contact uncertainties on system behavior, a comprehensive performance analysis is conducted by utilizing a time- and frequency-domain approach, offering deep insights into how impedance modulation shapes both transient dynamics and steady-state accuracy across varying environmental conditions. A high-fidelity simulation platform based on MATLAB (version 2021b) multi-toolbox co-simulation is developed to emulate realistic robotic contact conditions. Quantitative results show that the IIC framework significantly reduces settling time, overshoot, and undershoot under dynamic contact conditions, while maintaining stability and generalization across a broad range of environments.

Keywords: intelligent impedance control (IIC); deep reinforcement learning (DRL); force–motion control of manipulator; unknown environment; adaptive control

Academic Editor: Xiong Luo

Received: 2 July 2025
Revised: 8 August 2025
Accepted: 9 August 2025
Published: 11 August 2025

Citation: Shao, H.; Hu, W.; Yang, L.; Wang, W.; Suzuki, S.; Gao, Z. Intelligent Impedance Strategy for Force–Motion Control of Robotic Manipulators in Unknown Environments via Expert-Guided Deep Reinforcement Learning. *Processes* **2025**, *13*, 2526. https://doi.org/10.3390/pr13082526

1. Introduction

The evolution of robotic systems from structured industrial settings to unstructured and unknown environments—ranging from collaborative manufacturing to minimally invasive surgery—has necessitated advanced control paradigms capable of reconciling motion precision with safe, adaptive physical interaction. Traditional force control methods, such as impedance control (IC) and hybrid position/force control, have enabled robots to interact with partially modeled environments well by regulating contact forces through

stiffness/damping adjustments. However, their fixed-parameter architectures are likely to fail when confronted with dynamic uncertainties inherent to real-world applications: composite materials with time-varying stiffness [1,2], collaborative tasks requiring human-like adaptability [3], or biological tissues with nonlinear viscoelasticity [4].

Adaptive impedance control (AIC) has emerged as a promising solution by allowing real-time tuning of impedance parameters—stiffness (K), damping (B), and inertia (M)—based on sensory feedback. Bridging classical impedance control with modern adaptive control theory [5–15], AIC improves robustness and adaptability in uncertain environments.

In the realm of model-based AIC (MBAIC), many notable contributions have been made. For example, regressor-based adaptive control for manipulator–environment dynamics and model reference adaptive control for human–robot interaction scenarios targeted structured uncertainties in [5–7], while regressor-free approaches employed function approximation techniques in [8,9] or disturbance observers in [10] to handle nonlinearities and unknown external inputs. Notably, ref. [9] addressed both model uncertainties and contact disturbances using universal approximators. More recently, adaptive laws with error iterative mechanisms in [11] dynamically adjusted damping parameters, yielding promising steady-state tracking results. Recent studies evolved to address contact discontinuities, time-varying parameters, and unstructured environment uncertainties. For example, in [12], adaptive force-based control was represented to manage unknown terrain reaction forces, payload variations, and unmodeled dynamics in real time. In [13], an iterative stiffness adaptive method was developed for unknown terrain stiffness and geometry. Additionally, finite-time backstepping in [14] handled time-varying disturbances, fast convergence, unknown environments, and abrupt force transitions. Ref. [15] handled human-induced uncertainties, time-varying motion intentions, and interaction force variations in assistive tasks.

Despite these advances, MBAIC methods still face challenges in highly unstructured or dynamically evolving environments due to their reliance on accurate system models and sensitivity to parameter uncertainty. In particular, achieving both robust transient response and steady-state accuracy under unknown, time-varying contact conditions remains a critical open problem. These limitations have prompted a growing shift toward data-driven approaches, such as reinforcement learning and deep neural networks, which offer real-time adaptability without explicit model knowledge. This fusion of adaptive control theory and machine learning is expected to underpin the next generation of robotic systems, enabling greater versatility in domains ranging from autonomous assembly [16] to limb rehabilitation [17].

Several studies have explored hybrid architectures that combine model-based control principles with data-driven learning techniques to enhance adaptation in unknown environments. For instance, trajectory learning via iterative control [18], reinforcement learning for endpoint tracking [19], and impedance adaptation using integral reinforcement learning [20] have shown promising results. Gradient-based impedance learning schemes were also explored in [21]. However, these methods typically address either trajectory or force adaptation in isolation, offer limited analysis of force control performance, and often require manual tuning or are restricted to specific tasks. To some extent, these issues hinder their generalization to diverse or unseen contact scenarios. To further enhance adaptability in both free and constrained motion tasks, ref. [22] introduced a fuzzy neural network within a broad learning framework, while [23] proposed a reference adaptation approach using trajectory parameterization and iterative learning to balance tracking accuracy and force minimization. Although effective, these works often rely on partially known dynamics or require explicit modeling of environmental stiffness and damping,

limiting their adaptability and maintaining a degree of model dependency. Meanwhile, recent reinforcement-learning-based approaches [24,25] show promise but generally demand extensive exploration or rely on predefined movement primitives to a certain degree, which can still hinder real-time adaptation and pose safety risks in physical human–robot interaction scenarios. More recently, ref. [26] developed a DRL-based adaptive variable impedance control strategy that ensures provable stability while improving sample efficiency, generalization capability, and robust force tracking under uncertain conditions. These recent advances demonstrate the potential of hybrid control architectures that combine the stability and interpretability of model-based approaches with the adaptability and flexibility of data-driven learning. By leveraging the strengths of both paradigms, such methods accelerate the realization of robust and intelligent robotic systems capable of operating effectively in complex, dynamic, and uncertain environments.

Building on the above analyses, this study seeks to enhance robotic force–motion control in unknown, time-varying contact environments, with a focus on accompanied force tracking and exploratory contact motions. Emphasis is placed on improving contact dynamics, steady state force–motion performance, and learning efficiency by reducing exploration demands. The main contributions of this work are as follows:

(i) To enable intelligent and robust impedance adaptation in unknown environments, we propose a reinforcement learning framework that integrates a model-based expert to mitigate key limitations of DRL in continuous control domains. Unlike conventional approaches, our core innovation lies in reformulating the decoupled force–position interaction as a continuous Markov Decision Process (MDP). Within this formulation, a Deep Deterministic Policy Gradient (DDPG) algorithm, combined with expert-guided exploration, optimizes impedance parameters through an actor–critic architecture by leveraging a tailored reward function that reflects impedance regulation objectives.

(ii) To accelerate policy convergence and ensure initial contact stability, a warm-start strategy is employed using behavior cloning from a conventional adaptive variable impedance control law. This expert controller, based on force error feedback, serves as a demonstrator that provides prior knowledge for early-stage DDPG training. As a result, our method significantly enhances learning efficiency, achieving a 37.5% faster convergence rate—requiring only 125 episodes compared to 200 in the vanilla DDPG baseline—for complex force–motion coordination tasks involving transitions from flat to curved surfaces.

(iii) A comprehensive time- and frequency-domain performance analysis is conducted to evaluate both the dynamic and steady-state behavior of the system under various sources of uncertainty, including variations in reference force, reference trajectory, and surface material and geometry. Quantitative results confirm that the theoretical analysis is in good agreement with the outcomes of simulation experiments. Moreover, the proposed IIC approach demonstrates strong generalization and robust stability across diverse simulated conditions, delivering superior performance in the presence of changing material properties, environmental stiffness, and external disturbances.

The remainder of this paper is organized as follows: In Section 2, the preliminaries are presented, including the dynamic modeling of the robot–environment interaction and the formulation of impedance control under uncertain conditions. Section 3 details the proposed intelligent impedance control strategy, which incorporates expert guidance with a reinforcement learning framework. In Section 4, a comprehensive performance evaluation is provided, analyzing both dynamic and steady-state behaviors through time- and frequency-domain methods. Section 5 demonstrates the effectiveness of the proposed approach through simulation studies under a high-fidelity IIC simulation framework with

various unknown task scenarios. Finally, Section 6 concludes the paper and indicates potential directions for future research.

2. Preliminaries

2.1. Dynamics of Robot–Environment Interaction

Consider the interaction between an n-degree-of-freedom (DOF) robotic manipulator and an unknown environment characterized by a profiled surface, as illustrated in Figure 1. Let $\{B\}$ and $\{T\}$ denote the coordinate frames attached to the robot base and the end-effector, respectively. The end-effector interacts with the environment under constraints imposed by surface geometry $\boldsymbol{X_e}$, contact forces, and time-varying environmental stiffness $\boldsymbol{K_e}$. $\boldsymbol{X}$ demonstrates the position of the manipulator end-effector. The dynamic model of the manipulator must therefore accommodate both internal actuator dynamics and external interaction forces, forming the foundation for impedance-based force control in uncertain and structured environments.

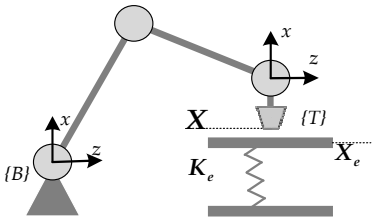

Figure 1. Interaction between robotic manipulator and environment.

In Cartesian space, considering a position inner loop regulated by calculated torque control as illustrated in Figure 2, the dynamics of the robotic manipulator under an impedance force control scheme with an admittance framework can be described as

$$M(q)\ddot{q} + C(q,\dot{q})\dot{q} + G(q) = \tau + \tau_d \tag{1}$$

where $q \in R^{n\times 1}$ is the joint position vector of the manipulator; $M(q) \in R^{n\times n}$ denotes the positive-definite inertia matrix; $C(q,\dot{q})\dot{q} \in R^{n\times 1}$ represents the Coriolis and centrifugal force vector, possessing the essential skew-symmetric property that preserves energy in Euler–Lagrange systems, i.e., $\dot{M}(q) - 2\dot{C}(q,\dot{q})$; $G(q) \in R^{n\times 1}$ is the gravitational torque vector arising from the manipulator's potential energy distribution; and $\tau \in R^{n\times 1}$ represents the control input torque vector, incorporating both commanded actuator effort and internal actuation dynamics. $\tau_d \in R^{n\times 1}$ embodies the lumped disturbance term aggregating exogenous perturbations from environmental interactions and endogenous uncertainties from model parameter variations; it will be compensated through the outer force control loop.

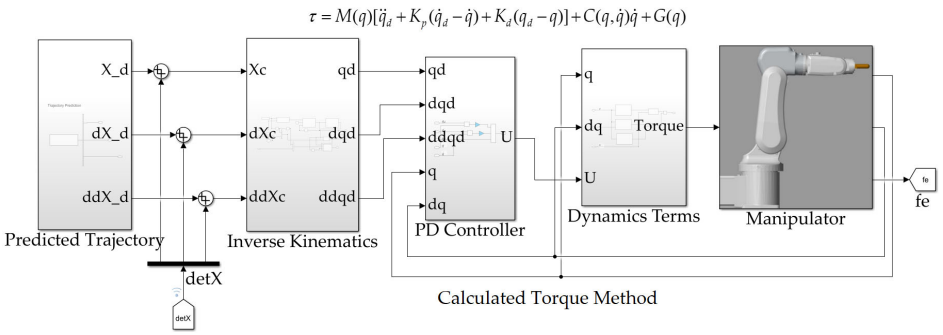

Figure 2. Position inner loop with calculated torque control.

Assumption 1. *The disturbance signal τ_d is piecewise differentiable and uniformly bounded by a constant, i.e., $\|\tau_d\| \leq \tau_{\max}, \forall t \geq 0$.*

To describe interaction with the environment, we consider a decoupled task-space representation, whereby operational axes are orthogonally decomposed into force-controlled directions (e.g., the normal direction z) and motion-controlled directions (e.g., tangential directions). In the force-controlled direction, quasi-static deformation due to contact dominates, allowing the environment to be approximated using a stiffness-dominant model. The resulting environment dynamics model can be described as

$$\boldsymbol{F_e} = \boldsymbol{K_e}(\boldsymbol{X_e} - \boldsymbol{X}) = \boldsymbol{K_e}\Delta\boldsymbol{X} \tag{2}$$

where $\boldsymbol{F_e} \in R^{m\times 1}$ denotes the external force vector at the manipulator's end-effector in Cartesian space determined by the number m of contact constraints imposed by the environment; $\boldsymbol{K_e} \in R^{m\times m}$ is the equivalent environmental stiffness matrix; $\boldsymbol{X} \in R^{m\times 1}$ is the measured end-effector pose; $\boldsymbol{X_e} \in R^{m\times 1}$ represents the nominal configuration of the environmental contact surface; and $\Delta\boldsymbol{X} \in R^{m\times 1}$ denotes the quasi-static deformation due to contact compliance.

This formulation serves as a simplified representation of environmental dynamics, forming the basis for designing an adaptive impedance controller and facilitating the analysis of associated uncertainties as follows.

2.2. Impedance Control in Uncertain Environments

From an engineering perspective, implementing force–motion control in unknown or unstructured environments presents persistent and multifaceted challenges. These include, but are not limited to, accurate environmental perception, precise dynamic modeling, robust control design, efficient learning, real-time safety assurance, and computational efficiency. Within this context, the desired impedance behavior between the manipulator and the environment can be represented by the following dynamic relationship:

$$\boldsymbol{M_d}\Delta\ddot{\boldsymbol{X}} + \boldsymbol{B_d}\Delta\dot{\boldsymbol{X}} + \boldsymbol{K_d}\Delta\boldsymbol{X} = \Delta\boldsymbol{F} \tag{3}$$

where $\boldsymbol{M_d}$, $\boldsymbol{B_d}$ and $\boldsymbol{K_d} \in R^{m\times m}$ are the desired inertia, damping, and stiffness matrices in Cartesian space, respectively; $\boldsymbol{\Delta F}=\boldsymbol{F_d} - \boldsymbol{F_e} \in R^{m\times 1}$ denotes the force deviation from the desired force $\boldsymbol{F_d}$ and the measured contact force $\boldsymbol{F_e}$. Under the admittance control framework, Equation (3) is used to derive the predicted environmental deformation $\boldsymbol{\Delta X}$ and desired trajectory $\boldsymbol{X_d}$ in response to force deviations $\boldsymbol{\Delta F}$.

When the actual position of the environment $\hat{\boldsymbol{X}}_e$ deviates from its nominal estimate, the difference can be captured by the tracking error $\delta\boldsymbol{X_e}$, which encapsulates the uncertainty in the environment's location and compliance. In this case, the expected quasi-static deformation of the environment can be approximated as

$$\boldsymbol{\Delta\hat{X}}=\boldsymbol{X}-\hat{\boldsymbol{X}}_e=\boldsymbol{X}-(\boldsymbol{X_e}+\delta\boldsymbol{X_e})=\boldsymbol{\Delta X}-\delta\boldsymbol{X_e} \tag{4}$$

Incorporating this uncertainty into the impedance model leads to a modified formulation:

$$\boldsymbol{M_d}\boldsymbol{\Delta}\ddot{\hat{\boldsymbol{X}}}+\boldsymbol{B_d}\boldsymbol{\Delta}\dot{\hat{\boldsymbol{X}}}+\boldsymbol{K_d}\boldsymbol{\Delta\hat{X}}=\boldsymbol{\Delta F}+\delta\boldsymbol{F} \tag{5}$$

where $\delta\boldsymbol{F}$ captures the force deviation arising from inner-loop control errors, unmodeled or time-varying environmental perturbations. Equation (5) reveals several embedded technical challenges—namely, environmental uncertainty, motion–force dynamic coupling, and the need for real-time environmental cognition to maintain control stability.

In practice, environmental dynamics are rarely static. They often exhibit time-varying or discontinuous properties—for instance, contact stiffness can span several orders of magnitude, ranging from soft biological tissues (~10 N/m) to rigid metallic surfaces (~10^6 N/m), while damping coefficients may fluctuate with material temperature or fatigue effects [27,28]. These changes significantly influence the manipulator's force control performance. Moreover, force–motion coupling in Cartesian space further complicates control, especially motion operations where inertial forces become dominant. This is particularly problematic for multi-degree-of-freedom (DoF) manipulators, where translational and rotational dynamics are often intricately intertwined [29]. As mentioned above in Section 1, some model-based adaptive impedance controllers based on Lyapunov redesign have demonstrated effectiveness under structured or unstructured conditions [8–12]. However, they tend to struggle with rapid environmental transitions. In contrast, reinforcement learning (RL) approaches, such as Q-learning, often suffer from poor convergence performance in continuous force control tasks, typically requiring over 10,000 training episodes to achieve acceptable policies [30].

Given the above challenges, it is evident that determining optimal impedance parameters and accurately estimating $\boldsymbol{\Delta}\hat{\boldsymbol{X}}$ with an admittance framework in real time under dynamic environmental uncertainties is nontrivial. Therefore, adaptive or learning-based impedance control strategies are indispensable for enabling robust and flexible interaction in unstructured and unpredictable environments.

3. Expert-Guided DRL for IIC Strategy

To combine the advantages of both traditional model-based and entirely learning-based approaches discussed in prior sections, we propose a position/force-decoupled intelligent impedance control framework incorporating a strategy network, as illustrated in Figure 3. This architecture achieves closed-loop adaptation of impedance parameters through trial-and-error interaction with various unknown environmental dynamics. The core innovation lies in the integration of a DRL policy with an expert-guided strategy network that makes it autonomously tune impedance parameters, thereby enabling robust and sample-efficient force tracking in environments characterized by time-varying and uncertain contact properties.

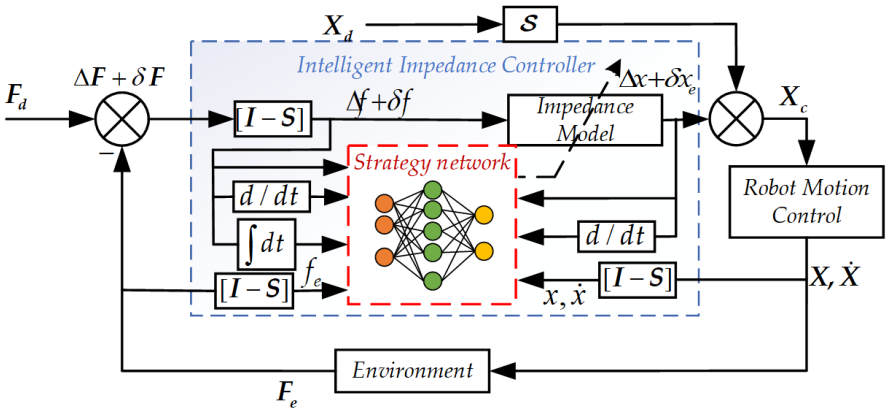

Figure 3. Block diagram of intelligent impedance control with hybrid control architecture.

To decouple motion control along the x and y directions from force control along the z-axis (the force-constrained direction), a selection matrix is introduced:

$$\boldsymbol{S} = diag[110] \tag{6}$$

This matrix facilitates orthogonal decomposition of the task space while preserving dynamic consistency. Focusing on the force-constrained direction allows the multidimensional impedance adaptation problem to be reduced to a one-dimensional formulation, simplifying controller training while retaining key system dynamics. In this simplified setting, Equation (5) can be reformulated as

$$m_0\Delta\ddot{\hat{x}} + (b_0 + \Delta b)\Delta\dot{\hat{x}} + (k_0 + \Delta k)\Delta\hat{x} = \Delta f + \delta f \tag{7}$$

where m_0, b_0, and k_0 represent the nominal inertia, damping, and stiffness parameters, and Δb and Δk are compensation terms from the adjustable impedance strategies $\mu^b = b_0 + \Delta b$ and $\mu^k = k_0 + \Delta k$ learned through the DRL agent under uncertain conditions. By substituting the predicted environmental deformation as expressed in Equation (4) into Equation (7), we obtain

$$m_0\Delta\ddot{x} + b_0\Delta\dot{x} + k_0\Delta x = \Delta f + (\delta f - m_0\delta\ddot{x}_e - b_0\delta\dot{x}_e) - (\Delta b\Delta\dot{x} + \Delta k\Delta x) = \Delta f + f_1 - f_2 \tag{8}$$

Here, $f_1 = \delta f - m_0\delta\ddot{x}_e - b_0\delta\dot{x}_e$ and $f_2 = \Delta b\Delta\dot{x} + \Delta k\Delta x$ denote the environmental uncertainties and the corresponding adjustable compensation term, respectively. As illustrated in Equation (8) and Figure 3, the impedance adaptation process can be achieved through the policy network that autonomously tunes the damping and stiffness parameters. This policy adjustment is performed by iterative trial-and-error exploration in response to force–motion tasks with the unknown interaction dynamics, using only the initial contact point as prior information. By doing so, the estimated force f_2 progressively converges to unknown dynamics f_1, effectively attenuating the influence of unmodeled environmental variations and dynamic disturbances. The closed-loop structure shown in Figure 3 also highlights the integration of real-time sensory feedback into the learning process, allowing the learned policy to adapt impedance parameters dynamically, enabling robust force tracking and adaptive and generalized behavior in complex and uncertain contact scenarios. To model dynamic agent–environment interactions for persistent contact operation, the robotic force control problem under environmental variability is formally framed as a Markov Decision Process (MDP), necessitating systematic design of state observables, action policies, and reward function topologies.

3.1. Design of State Space

To promote adaptive impedance control under unknown interaction dynamics, the state space is carefully constructed to capture the nonlinear coupling between the robot's end-effector motion and contact force behavior in uncertain environments. The state vector $\boldsymbol{s_t}$ at time t is defined as

$$\boldsymbol{s_t} = \left\{ f_{e,t}, \epsilon_f, \dot{\epsilon}_f, \int \epsilon_f, x_t, \dot{x}_t, \epsilon_x, \dot{\epsilon}_x \right\} \tag{9}$$

Here, $f_{e,t}$ denotes the contact force in the force-constrained direction. The vector $\left\{\epsilon_f, \dot{\epsilon}_f, \int \epsilon_f\right\}$ represents the force tracking error, its time derivative, and accumulated integral, forming a proportional–integral–derivative (PID)-like representation that enhances robustness to fluctuating force disturbances. The terms $\{x_t, \dot{x}_t\}$ correspond to the measured end-effector position and velocity in the same direction, while $\{\epsilon_x, \dot{\epsilon}_x\}$ capture the environment deformation and its rate of change.

This multidimensional state formulation enables the policy network to perceive and respond to key dynamic indicators, including real-time force deviation, motion tracking accuracy, and the system's transient behavior. By embedding both force regulation and kinematic feedback into the state space, this design ensures observability and supports

stable impedance adaptation in contact tasks involving time-varying, uncertain environments. It forms the foundation for real-time policy optimization after learning described in subsequent sections.

3.2. Design of Action Space

In the proposed IIC framework, the design of the action space plays a critical role in enabling the learning agent to effectively modulate the robot's dynamic behavior during interaction with unknown environments. Rather than directly commanding the joint torques or position trajectories, the action space $\boldsymbol{a_t}$ is defined in terms of the impedance parameters—specifically, the damping and stiffness coefficients within the force-constrained direction as follows:

$$\boldsymbol{a_t} = \left\{ \mu^b, \mu^k \right\} \in \boldsymbol{A} \tag{10}$$

where $\boldsymbol{A}$ denotes a bounded set of admissible impedance parameter variations. This formulation ensures that the learned policy adheres to the physical structure of impedance control, thereby maintaining system stability while allowing adaptive behavior.

To ensure both training efficiency and control safety, the action space $\boldsymbol{A}$ must be carefully constrained. If the permissible range is too narrow, the agent may be unable to explore effective parameter configurations. Conversely, an overly broad action space not only increases the risk of instability but also leads to inefficient exploration and elevated control energy consumption.

Remark 1. *The design of the action space is governed by two critical principles:*

(i) Stability Assurance: The bounds of $\boldsymbol{A}$ are derived through time- and frequency-domain stability analysis (see Section 4.2), ensuring that all learned impedance profiles yield asymptotic convergence in force tracking and maintain system passivity under nominal operating conditions.

(ii) Exploration Efficiency: Empirical evidence [31,32] shows that action spaces with excessively large ranges introduce high-dimensional exploration burdens [33,34] without proportionate gains in control performance. Hence, constraining $\boldsymbol{A}$ within theoretically justified and empirically effective limits is key to balancing learning efficiency with physical stability.

3.3. Reward Function

During each training episode, the robotic arm executes continuous force tracking tasks while interacting with unknown and possibly time-varying environmental surface profiles. The design of the reward function plays a pivotal role in shaping the learning trajectory of the agent and directly affects its ability to achieve robust and adaptive control. Traditional sparse-reward schemes—based on binary success or failure signals—often fail to provide sufficient learning gradients in complex contact-rich environments. This results in slow convergence and suboptimal control performance, particularly when precise force regulation and rapid adaptation are required simultaneously. On the other hand, excessively complex or over-engineered reward formulations can obscure the learning objective due to competing criteria and ill-defined trade-offs.

To address these limitations, the proposed reward function is explicitly designed to capture key performance objectives in uncertain environments, namely rapid dynamic response, minimal overshoot, and high steady-state tracking accuracy. The total reward is decomposed into multiple sub-components as follows:

$$r = w_1 r_d + w_2 r_s + w_3 r_t \tag{11}$$

where r_d is the dynamic reward, r_s represents steady-state reward, r_t denotes safety-based task termination reward. The weighting factors $w_1, w_2, w_3 \in R^+$ are hyperparameters

selected based on specific task requirements and system characteristics, balancing stability, responsiveness, and long-term precision.

3.3.1. Dynamic Reward r_d

The dynamic reward is specifically designed to encourage fast convergence and suppress force overshoot during the transient response phase. It leverages an exponential decay mechanism to penalize large deviations from the desired contact force, while partitioning the error space into intervals that reflect performance thresholds. The reward function is defined as

$$r_d = \sum_{i=1}^{m} R_i e^{-\alpha_i |f_d - f_{e,t}|}, |f_d - f_{e,t}| \in I_i \tag{12}$$

where m is the number of intervals dividing the error magnitude domain, f_d is the desired contact force, R_i is the reward coefficient for the i-th interval, α_i is the corresponding exponential decay rate, and I_i denotes the i-th partitioned region of the force deviation space.

This formulation allows dynamic sensitivity tuning across different error ranges. Larger errors invoke harsher penalties, encouraging the agent to swiftly reduce force deviation in early transient phases, while finer rewards in near-zero-error regions promote high precision during convergence. Compared to standard reward shaping methods, this interval-based decay approach provides several advantages:

- It enables task-specific sensitivity adjustment via interval-wise tuning;
- It offers robustness to sensor noise by grouping small deviations within thresholds, preventing reward instability due to measurement fluctuations [35];
- It supports generalization across surface types, as tuning the number of intervals and decay constants enables adaptation to various material properties and dynamics.

Overall, the dynamic reward incentivizes the agent to minimize contact force errors efficiently and stabilize interaction dynamics, thereby supporting both responsiveness and robustness in unpredictable environments.

3.3.2. Steady-State Reward r_s

The steady-state reward is designed to encourage precise force tracking by reinforcing small steady-state deviations and promoting consistent accuracy during sustained contact. It is mathematically defined as

$$r_s = -\log(|f_d - f_{e,t}| + \epsilon) \tag{13}$$

where ϵ is a small positive constant introduced to avoid singularity as the tracking error approaches zero. The use of a logarithmic function provides a continuous and smooth reward gradient, which increases sharply as the contact force error $|f_d - f_{e,t}|$ diminishes. This characteristic inherently motivates the agent to minimize long-term deviations and fine-tune its behavior around the desired force target.

The logarithmic reward structure also contributes to learning stability. Unlike linear or quadratic penalties, which may excessively punish large errors and destabilize learning, the logarithmic compression scales the penalty more moderately. This nonlinearity allows the agent to maintain learning progress even in the presence of occasional transient disturbances or sensing noise, making the training process more robust to outliers and irregularities in contact dynamics. As a result, it is expected that the steady-state reward supports both high-precision regulation and stable convergence in uncertain and variable environments.

3.3.3. Task Termination Reward r_t

The task termination reward incorporates a safety-oriented learning mechanism to ensure that the agent develops both stable and safe behaviors during force-interaction tasks. If any predefined safety constraint is violated during the training process, the episode is immediately terminated and penalized, effectively discouraging unsafe exploration and promoting policy robustness. Specifically, the agent's behavior is evaluated at each time step based on the following two critical safety conditions:

- Feasibility of the Position Command:

The commanded end-effector position $\boldsymbol{X_d}$ must correspond to a valid inverse kinematic solution, ensuring that the resulting joint configuration q remains within the allowable joint range $[q_{\min}, q_{\max}]$. This condition prevents physically infeasible or potentially damaging commands.

- Contact Force Constraint:

The actual contact force $f_{e,t}$ must remain below a predefined safe threshold $f_{e,\max}$, thereby avoiding excessive interaction forces that could compromise the mechanical integrity of the robot or its environment.

The task termination reward is defined as

$$r_t = \begin{cases} R_{success}, & \text{upon safe and successful task completion} \\ R_{failure}, & \text{upon failure due to safety violation or early termination} \end{cases} \quad (14)$$

where $R_{success} > 0$ is a fixed positive reward granted for safe task completion, and $R_{failure} < 0$ is a fixed negative penalty assigned upon safety violation or instability.

This reward component plays a dual role: it strengthens adherence to physical and safety constraints while guiding the agent toward reliable and responsible behavior in uncertain contact environments. By embedding safety directly into the reward structure, this mechanism enhances training stability and accelerates convergence toward viable control policies, especially in safety-critical or real-world applications [31,32,35].

3.4. DDPG with Expert Strategy

The strategy network architecture, detailed in Figure 4, is built upon an expert-guided Deep Deterministic Policy Gradient (DDPG) framework—a policy-based deep reinforcement learning algorithm tailored for continuous action spaces. By incorporating expert knowledge into the learning loop, this approach enhances stability and accelerates convergence, making it particularly suitable for addressing complex force control tasks in robotic arms interacting with uncertain and dynamic environments.

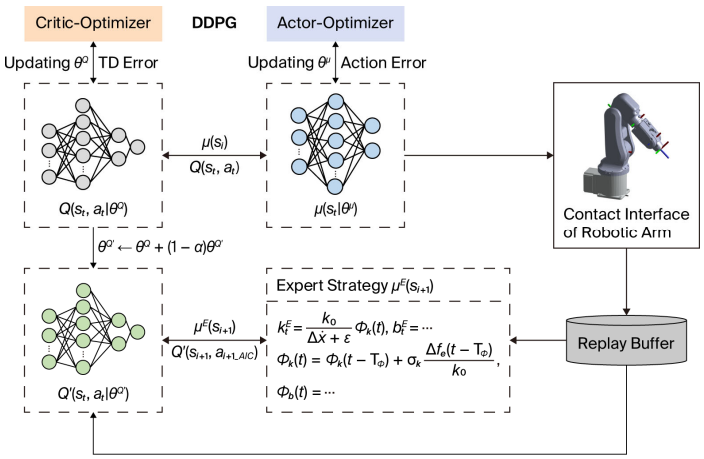

Figure 4. Expert-guided intelligent impedance control framework.

3.4.1. DDPG Algorithm [36,37]

As shown in Figure 4, the DDPG algorithm operates within an actor–critic architecture and is adapted here to support real-time impedance parameter adjustment. To improve sample efficiency and training stability, an experience replay buffer is employed to store state transitions collected from each complete contact episode during surface interaction. During each training iteration, a minibatch of N state transitions $(\boldsymbol{s_i}, \boldsymbol{a_i}, \boldsymbol{r_i}, \boldsymbol{s_{i+1}})$ is randomly sampled from this buffer to update the neural networks. Let $\boldsymbol{s_i}$, $\boldsymbol{a_i}$, and $\boldsymbol{r_i}$ denote the state, action, and reward at time step i, respectively, and let $\boldsymbol{s_{i+1}}$ represent the next state at time step $i + 1$. The critic network parameters are updated by minimizing the mean-squared Bellman error (MSBE) loss, which integrates two critical components: L2 regularization to prevent overfitting, and temporal-difference (TD) targets calculated via bootstrapping. Simultaneously, the actor network is updated by applying deterministic policy gradient ascent, where gradients are backpropagated to optimize the parameters of the current policy network. To balance exploration and exploitation in the continuous action space, the selected action is perturbed with Ornstein–Uhlenbeck (OU) noise. This temporally correlated noise facilitates smooth exploration trajectories, enhancing interaction efficiency in physical systems. To stabilize learning and reduce the risk of policy oscillation, target networks μ' and Q' are updated using a soft update mechanism by incrementally blending target network parameters with those of the current networks. This gradual synchronization technique prevents conflicting interactions between the actor's policy adjustments and the critic's value predictions, thereby stabilizing the learning process.

3.4.2. Expert Strategy

Despite its advantages, the DDPG algorithm faces persistent challenges in balancing exploration and exploitation of unknown dynamics, particularly during initial training stages when excessive random exploration substantially delays policy convergence. To mitigate this inefficiency in early-phase exploration, an expert-guided control strategy that enables accelerated acquisition of meaningful control experiences by providing structured prior knowledge to the agent is designed. This approach is especially pertinent in impedance control tasks, where conventional methods often struggle to adapt effectively to environmental uncertainties. To this end, our solution introduces an adaptive variable impedance law to obtain an expert-guided strategy, allowing the agent to rapidly acquire meaningful control behaviors without relying solely on gradient-based updates. By responding to dynamic interaction states with minimal computational overhead, the expert-guided process significantly accelerates the convergence of the DDPG learning algorithm.

The expert strategy iteratively refines the intelligent impedance strategy based on real-time tracking error feedback, starting from initially set impedance parameters derived from a weak-model-based approach. The stiffness and damping are adjusted according to both the rate of the error change and its accumulated history, enabling robust and stable adaptation under uncertain conditions. The actions selected by the expert strategy are defined as

$$\boldsymbol{a}_t^E = \mu^E(\boldsymbol{s}_t) = \left\{ b_t^E, k_t^E \right\} \tag{15}$$

where b_t^E and k_t^E represent the expert-determined adaptive stiffness and damping coefficients at time t. These coefficients are updated using the following iterative laws:

$$\begin{cases} k_t^E = \frac{k_0}{\Delta\dot{x}+\varepsilon}\Phi_k(t), \quad b_t^E = \frac{b_0}{\Delta\dot{x}+\varepsilon}\Phi_b(t) \\ \Phi_k(t) = \Phi_k(t - T_\Phi) + \sigma_k \frac{\Delta f_e(t-T_\Phi)}{k_0}, \\ \Phi_b(t) = \Phi_b(t - T_\Phi) + \sigma_b \frac{\Delta f_e(t-T_\Phi)}{b_0} \end{cases} \tag{16}$$

where T_Φ is the sampling interval, σ_k and σ_b are the update rates, and ε represents a small constant induced to prevent singularity at $\Delta\dot{x} = 0$ and provide numerical robustness.

During the expert-guided phase, the current policy network $\mu(\boldsymbol{s_t}|\boldsymbol{\theta}^\mu)$ is trained by minimizing the mean squared error between its output and the expert-generated strategy:

$$L(\theta^\mu) = \frac{1}{M}\sum_{i=1}^{M}\left\|\mu^E(s_i) - \mu(s_i \mid \theta^\mu)\right\|^2 \tag{17}$$

where M is the number of state transition samples randomly selected during the expert-guided period. Meanwhile, the critic network Q of the value function is updated using the standard DDPG loss based on the Bellman equation. During this phase, the target critic network Q' parameters are synchronized directly with those of the current critic to ensure consistency.

In this expert-guided learning period, the agent primarily mimics the behavior of the expert controller, allowing it to quickly learn feasible and stable policies. Once this phase concludes, the expert strategy μ^E is deactivated, and the strategy parameters of the agent will continue to be updated by employing the trained policy network, thereby preserving the architecture and dynamics of the original DDPG framework.

By embedding expert knowledge into the early training process, the agent benefits from a well-informed initialization. This pre-trained policy serves as a strong foundation for subsequent reinforcement learning, allowing the actor–critic architecture to progressively refine the control strategy. Ultimately, the agent acquires an optimal policy that exhibits both fast convergence and high adaptability in contact-rich and uncertain environments.

4. Performance Analysis

In this section, a time- and frequency-domain analysis approach is employed to evaluate the system's performance. As illustrated in Figure 5, a unidimensional force-constrained contact model is considered to assess the impact of the proposed intelligent impedance strategy on both the dynamic and steady-state behavior of force tracking under unknown environmental conditions. In this framework, the inner position control loop is approximated as a first-order inertial system with a time constant T, which includes position tracking error and modeling uncertainties. The outer impedance control loop modulates the desired force response based on this inner loop. Within the impedance control block, the parameters μ^b and μ^k represent the adjustable damping and stiffness coefficients, respectively, which are adaptively generated by the strategy network. This model allows for analytical insights into how the impedance adjustment influences system stability, transient response, and robustness in uncertain contact environments.

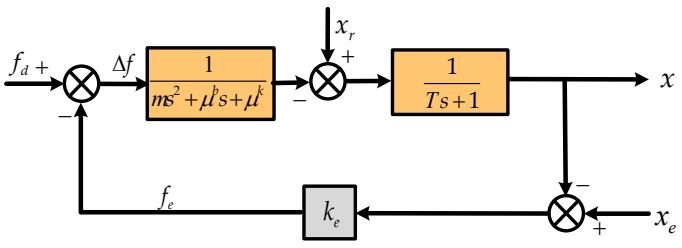

Figure 5. Simplified diagram of the force control outer loop.

4.1. Dynamics Performance Analysis

Considering learning convergent system, although stability analysis is not provided, methods such as the Routh–Hurwitz criterion remain theoretically applicable. To simplify the dynamic analysis, we assume the time constant is sufficiently small, allowing the first-

order term in Equation (22) to be neglected. Under this assumption, the system's damping ratio ζ and natural frequency ω_n can be expressed as

$$\zeta = \frac{\mu^b}{2\sqrt{m(\mu^k + k_e)}}, \omega_n = \sqrt{\frac{\mu^k + k_e}{m}} \tag{18}$$

This formulation reveals that the adjustable parameters μ^b and μ^k output by the impedance strategy network directly affect the system's damping and dynamic response characteristics. Specifically, increasing μ^b or decreasing μ^k enhances damping, thereby suppressing oscillatory behavior and reducing overshoot; adjusting ω_n by increasing μ^k modifies the system's transient response time. These findings indicate that in dynamic interaction scenarios involving rapidly varying or uncertain environments, the intelligent impedance controller should increase damping μ^b to improve vibration attenuation and tune ω_n to adapt the response rate according to contact dynamics. Moreover, in combination with the system's analysis in Equation (8), it becomes clear that optimal impedance parameter tuning effectively mitigates the destabilizing effects of environmental uncertainties, promoting robust and adaptive force control.

4.2. Steady-State Performance Analysis

It is assumed that the system is subject to modeling uncertainties from the position loop and external disturbances originating from three primary sources: the contact environment x_e, the reference trajectory x_r, and the reference force f_d. When the system remains stable, the corresponding error transfer functions for each component, denoted as $E_f(s)$, $E_{x_r}(s)$ and $E_{x_e}(s)$, can be derived as follows:

$$E_f(s) = \frac{\Delta F_1(s)}{F_d(s)} = \frac{(Ts+1)(ms^2 + \mu^b s + \mu^k)}{(Ts+1)(ms^2 + \mu^b s + \mu^k) + k_e} \tag{19}$$

$$E_{x_r}(s) = \frac{\Delta F_2(s)}{X_R(s)} = \frac{k_e(ms^2 + \mu^b s + \mu^k)}{(Ts+1)(ms^2 + \mu^b s + \mu^k) + k_e} \tag{20}$$

$$E_{x_e}(s) = \frac{\Delta F_3(s)}{X_E(s)} = \frac{-k_e(Ts+1)(ms^2 + \mu^b s + \mu^k)}{(Ts+1)(ms^2 + \mu^b s + \mu^k) + k_e} \tag{21}$$

Then, the lumped steady-state error (SSE) can be derived using the Final Value Theorem and the principle of linear superposition, as follows:

$$\epsilon_{ss} = \epsilon_{fss} + \epsilon_{x_r ss} + \epsilon_{x_e ss} = \lim_{s \to 0} s \cdot \Big(E_f(s)F_d(s) + E_{x_r}(s)X_R(s) + E_{x_e}(s)X_E(s) \Big) \tag{22}$$

Different contact surface profiles are detailed as follows:

(i) Planar Surface Contact

When the contact surface in the force-constrained direction is planar and the initial contact point can be treated as constant, force–motion control keeps a constant force tracking, and all inputs x_r, x_e, and f_d behave as step inputs. Let A_{f_d}, A_{x_r}, and A_{x_e} denote the amplitudes of these step signals. Under these conditions, SSE can be obtained as follows:

$$\epsilon_{ss} = \frac{\mu^k A_{f_d} + k_e \mu^k A_{x_r} - k_e \mu^k A_{x_e}}{\mu^k + k_e} \tag{23}$$

It is evident that $SSE \to 0$ as $\mu^k \to 0$. Thus, by regulating the impedance stiffness μ^k to zero, ideal steady-state force tracking performance can be achieved on a planar surface.

(ii) Sloped Surface Contact

For a sloped contact surface with an inclination angle, the SSE depends on the geometric profile of the environment. Besides the effects from f_d and x_r, the corresponding steady-state error expression of $\epsilon_{x_e ss}$ becomes

$$\epsilon_{x_e ss} = \lim_{s \to 0} s \cdot \frac{A_{x_e}}{s^2} \cdot E_3(s) = \begin{cases} -\frac{bA_{x_e}}{k_e} & \mu^k = 0 \\ -\infty & \mu^k \neq 0 \end{cases} \tag{24}$$

In this scenario, as long as the stiffness coefficient k remains nonzero, the cumulative tracking error increases progressively. Therefore, for optimal performance, the impedance strategy should dynamically reduce $\mu^k \to 0$ during interaction with inclined surfaces.

(iii) Sinusoidal Surface Contact

Since $\epsilon_{f ss}$ and $\epsilon_{x_r ss}$ remain unchanged, $\epsilon_{x_e ss}$ is analyzed as below. If the contact surface exhibits a sinusoidal profile, i.e., $x_e = A_{x_e} \sin \omega t$, the frequency-domain analysis is performed by substituting $s = j\omega$ into Equation (22). The resulting transfer function exhibits the following characteristics:

$$\epsilon_{x_e ss} = A_{x_e} A(\omega) \sin[\omega t + \varphi(\omega)] \tag{25}$$

$$\begin{aligned} A(\omega) &= k_e \sqrt{\frac{m^2T^2\omega^6 + (m^2 + \mu^{b2}T^2 - 2m\mu^k T^2)\omega^4 + (\mu^{b2} - 2m\mu^k + \mu^{k2}T^2)\omega^2 + \mu^{k2}}{m^2T^2\omega^6 + (m^2 + \mu^{b2}T^2 - 2m\mu^k T^2)\omega^4 + (\mu^{b2} - 2m\mu^k - 2mk_e - 2k_e\mu^b T + \mu^{k2}T^2)\omega^2 + (\mu^k + k_e)^2}} \\ and\ \varphi(\omega) &= \arctan\left(\frac{mT\omega^3 - \mu^k T\omega - \mu^b \omega}{\mu^k - \mu^b T\omega^2 - m\omega^2}\right) - \arctan\left(\frac{\mu^k T\omega + \mu^b \omega - mT\omega^3}{\mu^k + k_e - \mu^b T\omega^2 - m\omega^2}\right) \end{aligned} \tag{26}$$

The steady-state error contains a sinusoidal component at the same frequency ω as the environmental curvature, indicating that surface topology directly influences the periodic tracking error. Therefore, frequency-domain characteristics can be utilized to analyze how intelligent impedance regulation suppresses such errors.

5. Simulation Verification and Analysis

This section presents a high-fidelity IIC simulation framework developed within the MATLAB 2021b/Simulink environment, as depicted in Figure 6. To accurately replicate the dynamic behavior of the manipulator—including realistic contact interactions and control performance—a multi-toolbox co-simulation architecture is employed. Specifically, the virtual testbed is built using the Simscape Multibody physics simulation toolbox, which perfectly integrates the Unified Robot Description Format (URDF) model of the EFORT serial robotic manipulator as a benchmark robot system. The topological structure of this model is illustrated in Figure 6b.

The platform offers a robust and high-fidelity testbed for evaluating system-level performance under complex and uncertain conditions. The simulation experiments are designed to evaluate dynamic interactions with a single predefined touch-point under multi-source uncertainty, including variations in contact conditions and environmental compliance, all within a uniform motion trajectory. The simulation environment is parametrically aligned with the physical properties of the real system. Systematic comparisons are conducted against conventional IC and AIC as shown in Equation (16) to quantitatively evaluate performance in handling undetermined contact dynamics.

The adaptive impedance controller is trained using the Deep Learning Toolbox and Reinforcement Learning Toolbox, implementing the DDPG algorithm with expert-guided initialization. Training is conducted on a benchmarking platform featuring an Intel Core i7-7700K processor, utilizing the MATLAB Parallel Computing Toolbox in local worker mode to accelerate batch training and simulation cycles.

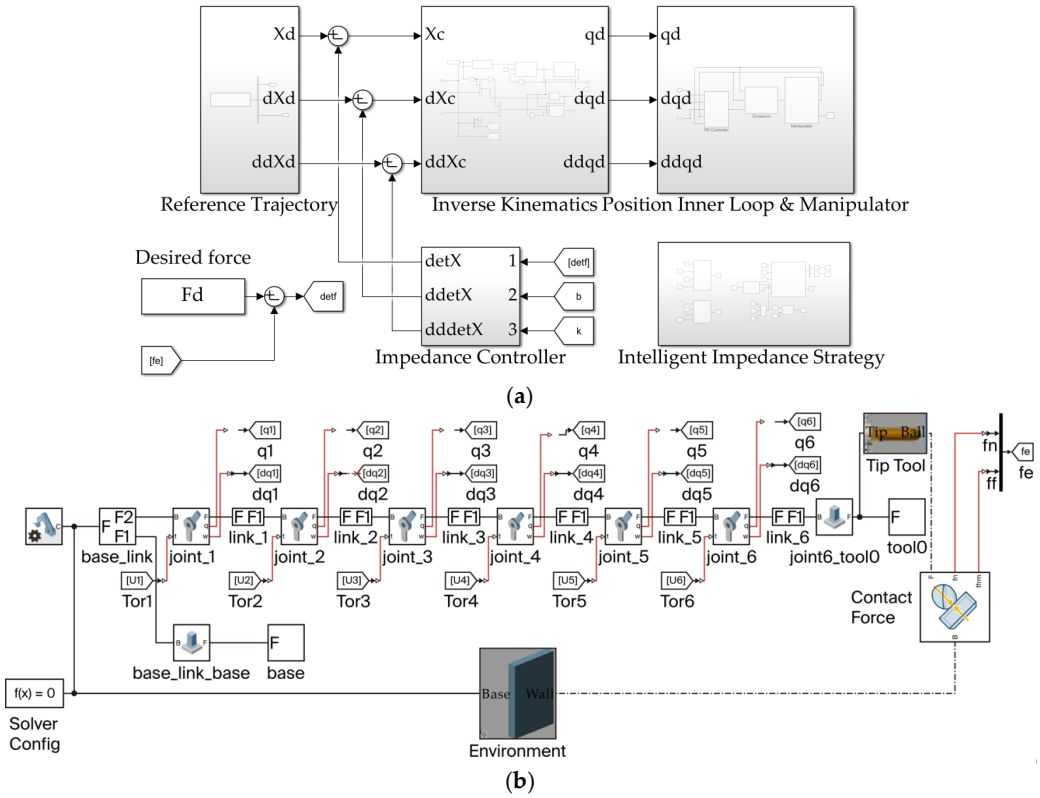

Figure 6. Schematic of IIC-based robotic manipulator contact platform. (**a**) MATLAB/Simulink high-fidelity framework; (**b**) topological configuration via Simscape Multibody.

5.1. Training Simulation

5.1.1. Learning Settings

The neural architectures underpinning the actor–critic framework, schematically illustrated in Figure 7, adopt distinct yet complementary feedforward designs. The actor network of policy π comprises two hidden layers, while the critic network of value function Q features a deeper structure with four hidden layers. Each layer contains 64 neurons, initialized with random parameters to encourage exploration in early training.

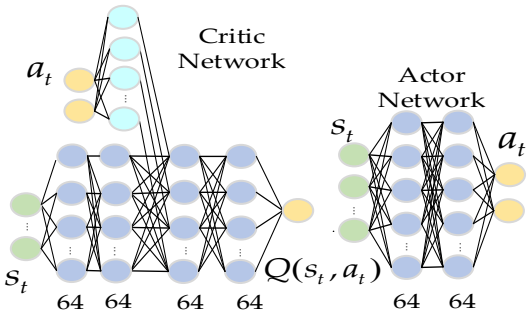

Figure 7. The actor–critic network framework.

The actor network receives an eight-dimensional state vector $\boldsymbol{s}_t$ as input, whereas the critic network adopts a skip connection structure. Specifically, the action vector $\boldsymbol{a}_t$ is element-wise summed with the output of the second hidden layer via trainable projection matrices. This residual integration mitigates gradient conflict during temporal-difference (TD) backpropagation, stabilizing convergence and improving learning robustness.

The output layers employ task-specific activation strategies to enforce meaningful constraints. The critic network uses rectified linear unit (ReLU) activations to ensure non-negative Q-value outputs, consistent with the Bellman equation's theoretical assumptions.

In contrast, the actor network applies a hyperbolic tangent (tanh) activation function followed by a scaling layer to constrain continuous control outputs within a predefined range. This bounded action representation not only reduces excessive control effort through saturation but also enhances policy generalization.

To prevent overfitting, L2 regularization with a scaling factor is applied, and all network parameters are optimized using the Adam optimizer, which improves convergence speed while reducing training latency and computational cost [37]. The specific training hyperparameters are summarized in Table 1. Each training episode corresponds to a 1.5 s interaction trajectory consisting of 1500 sampled transitions. These transitions are stored in a replay buffer of size 10^6. During training, a minibatch of 64 samples is drawn randomly from the buffer to update the actor and critic networks. Over the course of training 300 episodes with trial-and-error mode, approximately 450,000 unique transitions can be collected. Depending on the update frequency, once per time step, the agent performs up to 1500 gradient updates per episode. This cumulative data volume ensures sufficient experience diversity while maintaining the efficiency of the learning process.

Table 1. Network training parameters.

Parameters	Values
Sampling time/s	0.001
Episode length/s	1.5
Relay buffer length	10^6
Learning rate for critic	0.001
Learning rate for actor	0.0001
Discount factor γ	0.99
Random extract transitions N, M	64
Steps per episode H	1500
Total training episodes	300

5.1.2. Training Process

During training, the robotic arm autonomously explores unstructured contact environments through directionally guided interaction, executing adaptive contact motions at a constant velocity across unmodeled surface geometries. The simulation framework includes a variety of environmental profiles x_e, both planar and curved, to enhance generalization under force-controlled conditions, as illustrated in Figure 8. And the system predicts trajectory x_c in the force constraint direction by adjusting control actions accordingly.

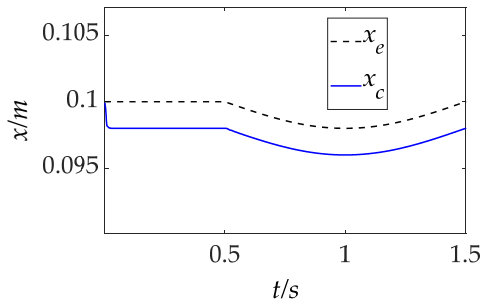

Figure 8. Trajectory variation in the direction of force constraint.

On the force control axis, the initial contact point with the environment is set at x_{e0}, with initial nominal impedance parameters defined as $m_0 = 2$ kg, $b_0 = 100$ Ns/m and $k_0 = 1000$ N/m. A change in the surface profile is introduced at t = 0.5 s to emulate dynamic environmental transitions. The environment stiffness is configured as $k_e = 5000$ N/m, and

the target contact force is maintained at $f_d = 10$ N, serving as the reference for force-tracking behavior throughout the training process.

The Intelligent Impedance controller Is trained under two distinct modalities: the baseline DDPG algorithm and the expert-integrated variant (EXPERT_DDPG). The comparative learning performance is visualized through the reward evolution profiles presented in Figure 9. The instantaneous reward trajectories for the IC, AIC, and IIC are computed using the policy optimization metric defined in Equation (11), while the average reward values of IIC are depicted as a statistical aggregation over 20 training rounds.

As shown in Figure 9a,b, the reward performance of traditional IC and AIC remains constant throughout the training process, as these controllers are not driven by reward-based optimization. In contrast, agents trained using the IIC framework exhibit a steadily increasing reward trend as the number of training episodes progresses. Although the reward curves display fluctuations due to exploration noise, both IIC variants eventually converge after approximately 200 and 125 training episodes, respectively. Notably, the EXPERT_DDPG algorithm achieves a 37.5% improvement in convergence efficiency over the baseline DDPG approach. These findings validate both the effectiveness and the enhanced training efficiency of the proposed IIC method cooperating with expert-guided strategies.

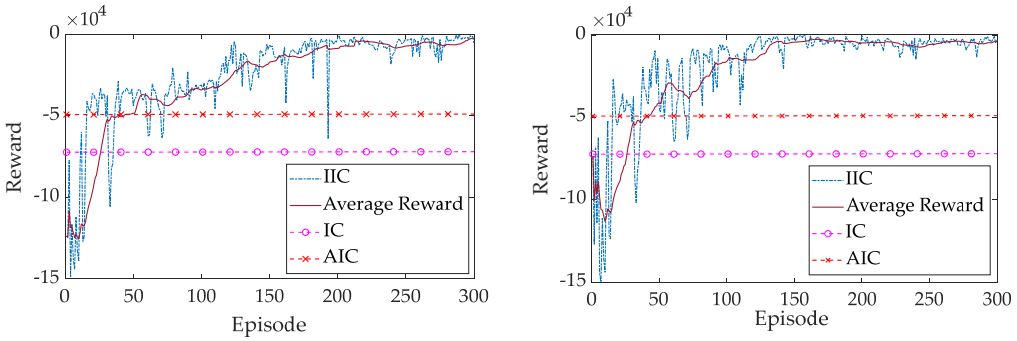

Figure 9. Reward curve. (**a**) Training process using DDPG; (**b**) training process using EXPERT_DDPG.

The Impedance parameter adjustment strategies learned by the IIC framework under both training paradigms are illustrated in Figure 10. In conjunction with the position trajectory variation depicted in Figure 8, it is evident that the action parameters undergo rapid changes when the robotic arm transitions into contact with the environment or encounters abrupt variations in surface profile. Notably, during step changes in the reference force, the stiffness μ^k and damping coefficient μ^b increase under the proposed strategy, effectively suppressing overshoot and improving response, which aligns well with the dynamic analysis presented in Section 4.2. These behaviors indicate that during the transient phase of force tracking control, the IIC framework successfully generates impedance control strategies that strike a balance between rapid response and effective damping. Consequently, the system achieves optimal dynamic performance throughout the interaction.

As the robotic arm transitions into steady-state contact, the impedance control strategy $\begin{Bmatrix} \mu^k & \mu^b \end{Bmatrix}$ produced by the IIC method continuously fine-tunes the action parameters to minimize tracking error within a narrow margin. This adaptive behavior is consistent with the learning objectives outlined in Section 2.2, demonstrating the framework's ability to sustain high-precision control in uncertain and dynamic environments.

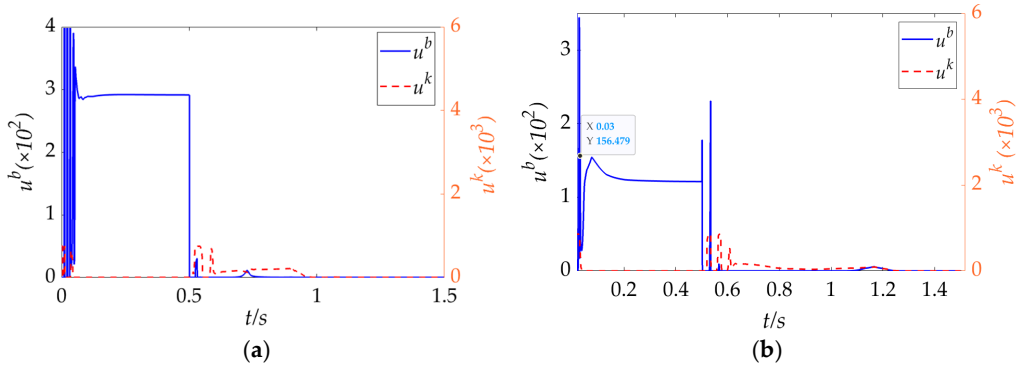

Figure 10. IIC strategies under two training modalities. (**a**) The IIC strategy using DDPG; (**b**) the IIC strategy using EXPERT_DDPG.

5.2. Comparative Evaluation in Uncertain Scenarios

To rigorously evaluate the force control performance of the IIC framework in diverse uncertain environments, a series of comparative experiments is conducted against IC and AIC methods. In these comparisons, IC uses appropriately selected fixed impedance parameters $m_0 = 2$ kg, $b_0 = 200$ Ns/m, and $k_0 = 0$ N/m, while AIC adopts the initial values described in Section 5.1.2 along with $k_0 = 0$ N/m. Additionally, the parameters σ_k and σ_b in Equation (16) are both set to 0.05.

The simulations encompass five representative uncertain contact scenarios, designed to reflect realistic and challenging operational conditions: (i) time-varying reference force, (ii) variable environmental stiffness, (iii) sloped surface contact, (iv) curved surface contact, and (v) contact environments subjected to external disturbances. These scenarios collectively test the adaptability, robustness, and dynamic response of each control strategy under dynamically changing conditions.

5.2.1. Time-Varying Reference Force

To evaluate the robustness of the IIC framework under dynamic task requirements, a basic planar contact environment is established, where the desired contact force in the force control direction varies over time as follows:

$$f_d = \begin{cases} 7\text{ N}, 0\text{ s} \leq t < 0.5\text{ s} \\ 11\text{ N}, 0.5\text{ s} \leq t < 1\text{ s} \\ 8\text{ N}, 1\text{ s} \leq t \leq 1.5\text{ s} \end{cases}$$

The resulting force tracking performance is illustrated in Figure 11. As shown, the conventional IC method exhibits relatively large tracking errors throughout the task. The AIC method, while maintaining good steady-state accuracy, suffers from pronounced overshoots and prolonged settling times following each step change in the reference force. In contrast, the proposed IIC approach achieves marked improvements in both transient and steady-state behaviors. Specifically, the IIC method demonstrates minimal overshoot and significantly reduced tracking error under the same conditions, thereby offering superior overall control performance.

Quantitative control metrics including initial overshoot, settling time, undershoot during falling transitions, and steady-state error (SSE) are summarized in Table 2. Compared to the baseline IC method, the IIC framework reduces overshoot and settling time by 74% and 44%, respectively, and reduces undershoot by 28%, while maintaining an SSE on the order of $O(10^{-2})$. These results confirm that the IIC method consistently delivers enhanced dynamic responsiveness and tracking precision for both rising and falling reference forces.

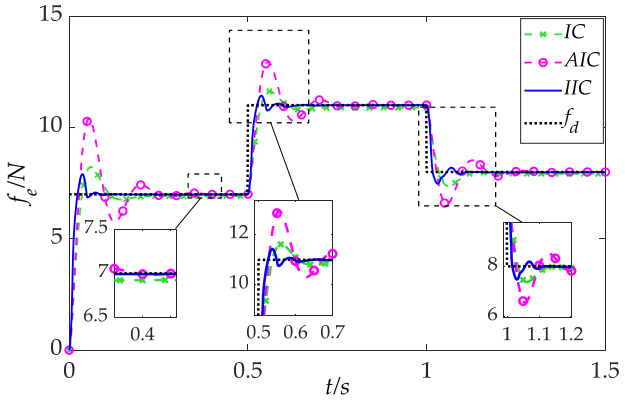

Figure 11. Force tracking with time-varying reference force.

Table 2. Performance metric comparison with time-varying reference force.

Method	Overshoot/%	Settling Time/s (5%)	SSE/N	Undershoot/%
IC	17.4	0.16	0.121	9.1
AIC	48.6	0.27	0.011	18.3
IIC	12.9	0.07	0.012	6.6

5.2.2. Variable Environmental Stiffness

In this scenario, the end-effector contact configuration of the robotic arm follows the setup described in Section 5.2.1, with the reference contact force maintained at a constant value. To quantitatively evaluate the proposed controller's adaptability to environmental stiffness variations, simulations are conducted to emulate real-world scenarios where stiffness changes abruptly through successive step transitions. The time-varying stiffness is defined as

$$k_e = \begin{cases} 3500\ \text{N/m}, 0\ \text{s} \leq t < 0.5\ \text{s} \\ 5000\ \text{N/m}, 0.5\ \text{s} \leq t < 1\ \text{s} \\ 7000\ \text{N/m}, 1\ \text{s} \leq t \leq 1.5\ \text{s} \end{cases}$$

The resulting force tracking responses for the IC, AIC, and IIC methods are presented in Figure 12. As illustrated, abrupt changes in environmental stiffness lead to significant degradation in the AIC method, with prolonged adjustment times and substantial oscillations characterized by longer settling times and larger undershoots. In contrast, although the peak oscillation amplitude remains roughly constant across methods, the IIC controller demonstrates rapidly enhanced adaptability to stiffness uncertainties. Following the first stiffness change, the IIC method reduces undershoot by 12% and settling time by 48% compared to the IC baseline. Similar improvements are observed after the second increase in stiffness. Overall, the IIC approach outperforms both IC and AIC by achieving faster response times and maintaining stable force tracking with minimal undershoot across all transitions. Detailed performance comparisons are summarized in Table 3.

Table 3. Performance comparison with changes in environmental stiffness.

Method	Undershoot/%	Settling Time/s (5%)	SSE/N
IC	8.8	0.11	0.155
AIC	20.1	0.21	0.008
IIC	7.4	0.053	0.009

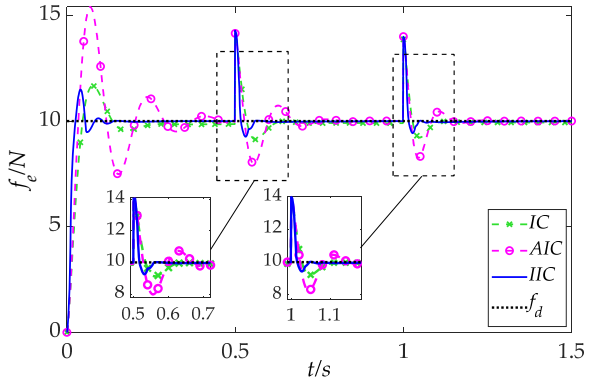

Figure 12. Force tracking with dynamic changes in environmental stiffness.

These results confirm that the IIC method offers superior robustness to dynamic variations in environmental stiffness and consistently delivers high-precision force tracking performance under these uncertain contact conditions.

5.2.3. Sloped Surface Contact

To assess the adaptability of the IIC method to dynamic positional variations in the environment, the robotic arm is tasked with operating in a composite terrain comprising both flat and sloped surfaces. The corresponding trajectory variations along the force control axis are illustrated in Figure 13.

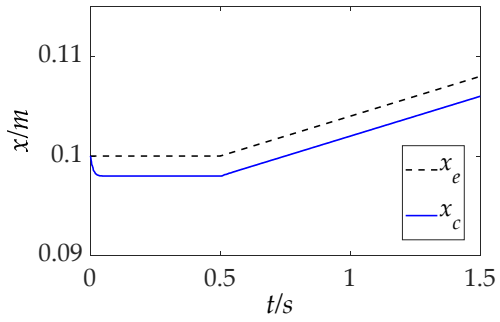

Figure 13. Trajectory variations with a sloped surface.

Force tracking results for this scenario are shown in Figure 14, with key performance metrics including overshoot, settling time, and SSE summarized in Table 4, aligned with changes in surface profile.

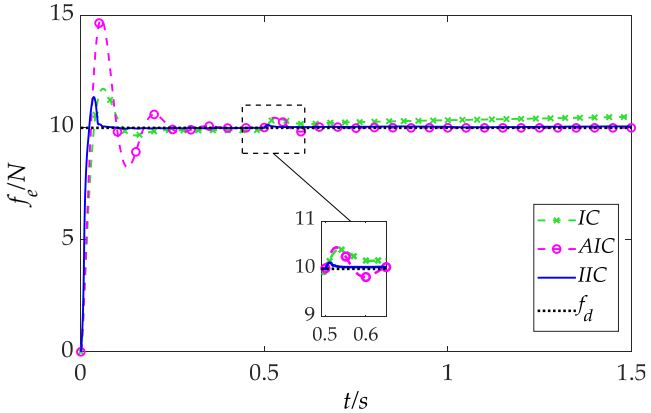

Figure 14. Force tracking on a sloped surface.

Table 4. Performance comparison on a slope.

Method	Overshoot/%	Settling Time/s (5%)	SSE/N
IC	4.2	0.09	0.491
AIC	4.4	0.1	0.011
IIC	1.2	0	0.058

As the robotic arm transitions from a flat to an inclined region, the IC method exhibits a gradual increase in force tracking error, consistent with the SSE analysis presented in Section 4.1 (ii). When a sudden geometric variation occurs at t = 0.5 s, the IC and AIC methods display similar dynamic performance, characterized by overshoot and prolonged settling time. In contrast, the IIC method maintains significantly improved tracking performance throughout the transition, exhibiting almost negligible minimal oscillation and relatively low SSE even when subjected to abrupt changes in surface inclination.

These results demonstrate the IIC controller's strong ability to adapt to positional uncertainties δx_e induced by inclined geometries, validating its robustness and effectiveness in unstructured or previously unknown sloped environments.

5.2.4. Curved Surface Contact

As in the training scenario described in Section 5.1.2, another representative complex surface composed of both flat and curved segments is introduced. The manipulator's trajectory variations along the normal contact direction in this environment are shown in Figure 15. The corresponding force tracking responses using the IC, AIC, and IIC methods are presented in Figure 16.

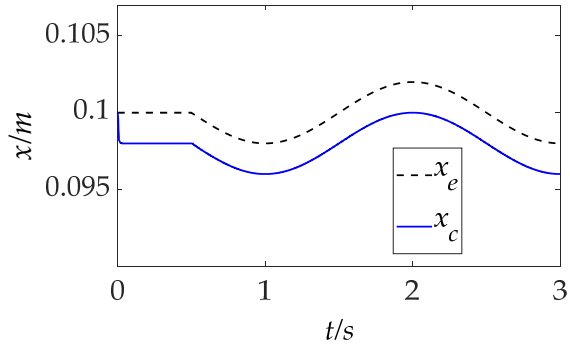

Figure 15. Trajectory variations with the curved surface.

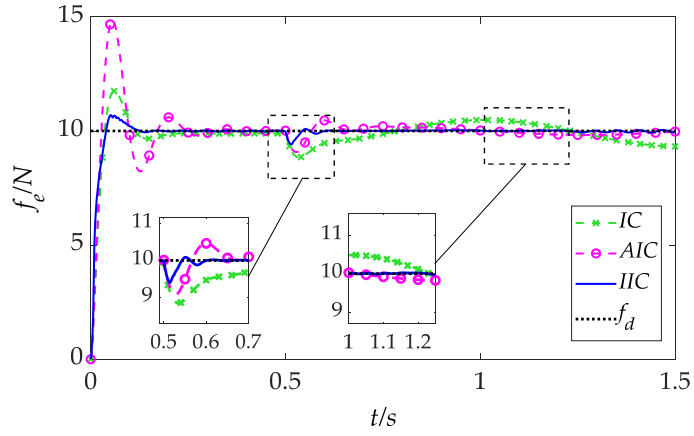

Figure 16. Force tracking performance on the curved surface.

As shown, when the environment undergoes positional variations due to the curved geometry, initial force oscillations manifest as undershoot caused by the concave contact region. The IC method exhibits sinusoidal-like fluctuations in tracking error, remaining within 5% but failing to reach within 2% tolerance. Although the AIC method achieves a relatively low SSE, it suffers from pronounced undershoot and oscillatory behavior, indicating less favorable dynamic characteristics. In contrast, the IIC method outperforms both baselines in terms of dynamic response and steady-state accuracy. It effectively suppresses overshoot and undershoot while improving overall tracking precision, as corroborated by the quantitative results in Table 5. Notably, at t = 0.5 s, when positional uncertainty is introduced by the curved surface, the IIC method achieves the smallest oscillation amplitude, reducing undershoot by 50% compared to the IC method through adaptive impedance adjustment. These results further validate the robustness and effectiveness of the IIC framework in handling dynamically varying contact conditions on complex curved surfaces.

Table 5. Performance comparison with the curved surface.

Method	Undershoot/%	Settling Time/s (2%)	SSE/N
IC	11.5	∞	0.481
AIC	9.2	0.134	0.175
IIC	5.7	0.033	0.062

5.2.5. Disturbance Rejection Under External Perturbations

To further evaluate the system's robustness against external force disturbances, a transient external perturbation with a specified amplitude of 3 N is applied in the normal direction between 0.5 s and 1.0 s. During this period, the robotic arm maintains constant motion along a flat surface while sustaining a target contact force under the same environmental stiffness conditions as described in Section 5.2.1. The resulting force tracking responses using IC, AIC, and IIC methods are shown in Figure 17, with quantitative performance metrics summarized in Table 6.

Table 6. Performance comparison of force tracking robustness.

Method	Undershoot/%	Settling Time (2%)	SSE/N
IC	5.8	0.09	0.11
AIC	14.5	0.158	0.021
IIC	5.4	0.053	0.012

Simulation results indicate that although all control strategies suppress the peak disturbance to a similar amplitude, the IIC framework exhibits superior dynamic convergence properties, achieving the smallest undershoot, shortest settling time, and lowest SSE in the presence of disturbances. Specifically, it achieves faster stabilization with a 41% reduction in settling time compared to the conventional IC method under identical disturbance conditions.

It is important to note that the contact dynamics with uncertainties, as described by Equation (5), pose significant challenges in simultaneously achieving satisfactory transient response and steady state performance to robustness. Any variation δf in operating conditions can lead to substantial parameter uncertainties, which represent a major control challenge. Nevertheless, the IIC method demonstrates strong disturbance rejection capability in the presence of force uncertainties induced by environmental perturbations. This effectively verifies the robustness and adaptability of the IIC approach in unknown and unstructured environments.

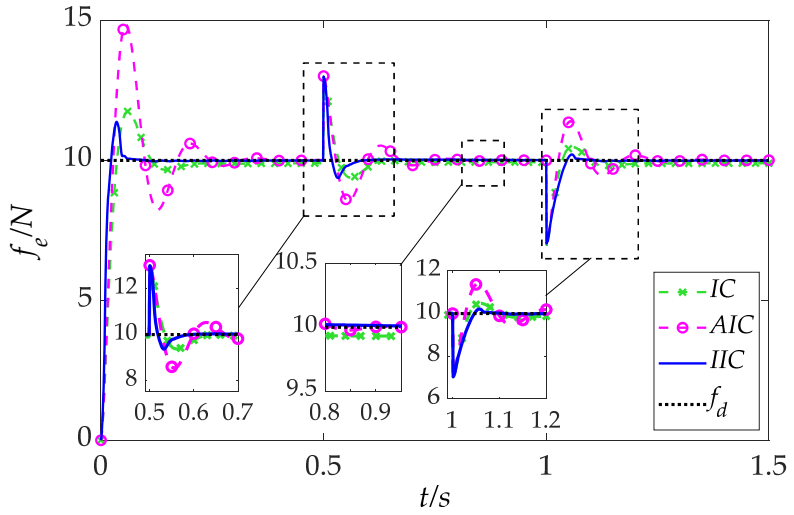

Figure 17. Force tracking performance with external disturbances.

6. Conclusions and Future Scope

6.1. Conclusions

This study proposed an Intelligent Impedance Control (IIC) framework that integrates model-based insights with a data-driven adaptive strategy using Deep Reinforcement Learning (DRL) to address the challenges of force control in robotic manipulators operating under uncertain environmental conditions. By formulating the control problem as a Markov Decision Process (MDP) and employing the Deep Deterministic Policy Gradient (DDPG) algorithm, the proposed method enables the autonomous learning of impedance control policies that adapt to a wide range of contact scenarios.

To enhance training efficiency, particularly during the early learning phase, an expert-guided adaptive impedance strategy with a conventional error-feedback iterative updating mechanism was introduced. This pre-trained policy effectively accelerates convergence and steers the learning process toward more stable and efficient control policies.

Comprehensive simulations conducted across a variety of uncertain contact environments—characterized by diverse surface profiles, varying material stiffness, and transient disturbances—demonstrated the superior generalization and adaptability of the IIC approach. Compared to conventional impedance control (IC) and adaptive impedance control (AIC), the IIC method consistently outperformed in terms of reduced overshoot, faster settling time, and lower steady-state error under unknown and complex environmental conditions.

6.2. Future Scope

The current evaluation is conducted entirely in high-fidelity simulation, which allows contact task adaptation within a certain range of environmental dynamic changes, especially where robot and environment dynamics can be realistically captured. These findings highlight the potential of DRL-based impedance control strategies in improving the robustness and adaptability of robotic manipulators in dynamic and unstructured environments.

Nevertheless, while the proposed hybrid adaptive impedance strategy demonstrates a degree of generalization capabilities, further research is needed to enhance its applicability to more complex and uncertain contact scenarios. Future studies could explore environments with greater variability, such as pronounced surface concavity, steeper inclination angles, and a broader spectrum of material stiffness. In addition, the real-world deployment may introduce additional factors such as sensor noise, complex unmodeled dynamics, and hardware constraints. Therefore, future work could focus on transferring the learned policy

to real hardware, potentially using sim-to-real techniques such as domain randomization or fine-tuning with real-world data.

Furthermore, we plan to incorporate formal verification methods by integrating model-based reasoning with data-driven learning to improve the interpretability, stability, and reliability of control strategies in unstructured, time-varying environments. Comprehensive ablation studies will also be conducted to isolate and quantify the contributions of key components, including the impedance regulation, expert guidance, and the DRL framework itself.

While integrating DRL, expert guidance for impedance control may increase system complexity, challenging real-time implementation. Modular design, adaptability–real-time balance, computational advances, and dynamic environment demands jointly create compelling research opportunities worthy of exploration. Ultimately, we hope our efforts can provide a solid foundation for robust policy learning and contribute to advancing sim-to-real transfer for complex robotic manipulation tasks.

Author Contributions: Conceptualization, H.S. and W.H.; methodology, H.S. and L.Y.; software, H.S., W.H. and L.Y.; validation, W.H.; investigation, H.S.; resources, W.H., W.W., S.S. and Z.G.; data curation, H.S. and L.Y.; writing—original draft preparation, H.S.; writing—review and editing, H.S., W.W., S.S. and Z.G.; supervision, H.S., W.W., S.S. and Z.G.; project administration, H.S.; funding acquisition, H.S., W.H. and Z.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Science Foundation of Fujian Province, China, grant number 2021J01291.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author(s).

Acknowledgments: The authors would like to acknowledge EFORT Intelligent Robot Co., Ltd., CN; Chiba University, JP; and Northumbria University, UK.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Liang, L.; Chen, Y.; Liao, L.; Sun, H.; Liu, Y. A novel impedance control method of rubber unstacking robot dealing with unpredictable and time-variable adhesion force. *Robot. Comput.-Integr. Manuf.* **2021**, *67*, 102038. [CrossRef]
2. Zeng, X.; Zhu, G.; Gao, Z.; Ji, R.; Ansari, J.; Lu, C. Surface polishing by industrial robots: A review. *Int. J. Adv. Manuf. Technol.* **2023**, *125*, 3981–4012. [CrossRef]
3. Ji, W.; Tang, C.; Xu, B.; He, G. Contact force modeling and variable damping impedance control of apple harvesting robot. *Comput. Electron. Agric.* **2022**, *198*, 107026. [CrossRef]
4. Famaey, N.; Sloten, J.V. Soft tissue modelling for applications in virtual surgery and surgical robotics. *Comput. Methods Biomech. Biomed. Eng.* **2008**, *11*, 351–366. [CrossRef]
5. Slotine, J.-J.E.; Li, W. On the Adaptive Control of Robot Manipulators. *Int. J. Robot. Res.* **1987**, *6*, 49–59. [CrossRef]
6. Kelly, R.; Carelli, R.; Amestegui, M.; Ortega, R. On Adaptive Impedance Control of Robot Manipulators. In Proceedings of the International Conference on Robotics and Automation, Scottsdale, AZ, USA, 14–19 May 1989; pp. 572–577. [CrossRef]
7. Sharifi, M.; Behzadipour, S.; Vossoughi, G.R. Nonlinear model reference adaptive impedance control for human–robot interactions. *Control Eng. Pract.* **2014**, *32*, 9–27. [CrossRef]
8. Chien, M.; Huang, A. Adaptive Impedance Control of Robot Manipulators based on Function Approximation Technique. *Robotica* **2004**, *22*, 395–403. [CrossRef]
9. Izadbakhsh, A.; Khorashadizadeh, S.; Ghandali, S. Robust adaptive impedance control of robot manipulators using Szász–Mirakyan operator as universal approximator. *ISA Trans.* **2020**, *106*, 1–11. [CrossRef]
10. Dong, Y.; Ren, B. UDE-Based Variable Impedance Control of Uncertain Robot Systems. *IEEE Trans. Syst. Man Cybern. Syst.* **2019**, *49*, 2487–2498. [CrossRef]
11. Duan, J.; Gan, Y.; Chen, M.; Dai, X. Adaptive variable impedance control for dynamic contact force tracking in uncertain environment. *Robot. Auton. Syst.* **2018**, *102*, 54–65. [CrossRef]

12. Xu, K.; Wang, S.; Yue, B.; Wang, J.; Peng, H.; Liu, D.; Chen, Z.; Shi, M. Adaptive impedance control with variable target stiffness for wheel-legged robot on complex unknown terrain. *Mechatronics* **2020**, *69*, 102388. [CrossRef]
13. Sombolestan, M.; Nguyen, Q. Adaptive Force-Based Control of Dynamic Legged Locomotion over Uneven Terrain. *IEEE Trans. Robot.* **2024**, *40*, 2462–2477. [CrossRef]
14. Ahmadi, S.M.; Taghadosi, M.B.; Nazmara, G. Adaptive finite-time impedance backstepping control for uncertain robotic systems interacting with unknown environments. *Int. J. Control* **2023**, *96*, 2671–2682. [CrossRef]
15. Ma, J.; Chen, H.; Liu, X.; Yang, Y.; Huang, D. Adaptive Impedance Control of a Human–Robotic System Based on Motion Intention Estimation and Output Constraints. *Appl. Sci.* **2025**, *15*, 1271. [CrossRef]
16. Zhou, Z.; Yang, X.; Zhang, X. Variable impedance control on contact-rich manipulation of a collaborative industrial mobile manipulator: An imitation learning approach. *Robot. Comput.-Integr. Manuf.* **2025**, *92*, 102896. [CrossRef]
17. Jiang, Z.; Wang, Z.; Lv, Q.; Yang, J. Impedance Learning-Based Hybrid Adaptive Control of Upper Limb Rehabilitation Robots. *Actuators* **2024**, *13*, 220. [CrossRef]
18. Xu, Q.; Sun, X. Adaptive Impedance Control of Robots with Reference Trajectory Learning. *IEEE Access* **2020**, *8*, 104967–104976. [CrossRef]
19. Xing, L.; Shuzhi, S.G.; Fei, Z.; Xuesong, M. Optimized Impedance Adaptation of Robot Manipulator Interacting with Unknown Environment. *IEEE Trans. Control Syst. Technol.* **2021**, *29*, 411–419. [CrossRef]
20. Peng, G.; Yang, C.; Li, Y.; Chen, C.L.P. Impedance and Trajectory Adaptation for Contact Robots Using Integral Reinforcement Learning. In Proceedings of the 37th Youth Academic Annual Conference of Chinese Association of Automation (YAC), Beijing, China, 11–13 November 2022; pp. 1542–1547. [CrossRef]
21. Li, Y.; Ge, S. Impedance Learning for Robots Interacting With Unknown Environments. *IEEE Trans. Control Syst. Technol.* **2014**, *22*, 1422–1432. [CrossRef]
22. Huang, H.; Yang, C.; Chen, C.L.P. Optimal Robot–Environment Interaction under Broad Fuzzy Neural Adaptive Control. *IEEE Trans. Cybern.* **2020**, *51*, 3824–3835. [CrossRef]
23. Wang, C.; Li, Y.; Ge, S.S.; Tong, H.L. Reference Adaptation for Robots in Physical Interactions with Unknown Environments. *IEEE Trans. Cybern.* **2017**, *47*, 3504–3515. [CrossRef]
24. Chang, C.; Haninger, K.; Shi, Y.; Yuan, C.; Chen, Z.; Zhang, J. Impedance Adaptation by Reinforcement Learning with Contact Dynamic Movement Primitives. *arXiv* **2022**, arXiv:2203.07191. [CrossRef]
25. Xing, X.; Burdet, E.; Si, W.; Yang, C.; Li, Y. Impedance learning for human-guided robots in contact with unknown environments. *IEEE Trans. Robot.* **2023**, *39*, 3719–3735. [CrossRef]
26. Li, Y.; Zeng, L.; Wang, Y.; Dong, E.; Zhang, S. Impedance Learning-Based Adaptive Force Tracking for Robot on Unknown Terrains. *IEEE Trans. Robot.* **2025**, *41*, 1404–1420. [CrossRef]
27. Huang, W.; Yu, G.; Xu, W.; Zhou, R. A Stochastic Dynamics Method for Time-Varying Damping Depending on Temperature/Frequency for Several Alloy Materials. *Materials* **2024**, *17*, 1207. [CrossRef]
28. Rosales, A.; Freidovich, L. Estimation of Time-Varying Parameters Defining Contact of a Planar Manipulator with a Surface. In Proceedings of the IEEE 62nd Conference on Decision and Control (CDC), Singapore, 13–15 December 2023; pp. 1392–1397. [CrossRef]
29. Pan, G.; Jia, Q.; Chen, G.; Wang, Y.; Sun, F. Analysis and Optimization of Motion Coupling for the Coordinated Operation of Flexible Multi-Arm Space Robots. *Actuators* **2023**, *12*, 198. [CrossRef]
30. Seyde, T.; Werner, P.; Schwarting, W.; Gilitschenski, I.; Riedmiller, M.A.; Rus, D.; Wulfmeier, M. Solving Continuous Control via Q-learning. *arXiv* **2022**, arXiv:2210.12566.
31. Varin, P.; Grossman, L.; Kuindersma, S. A Comparison of Action Spaces for Learning Manipulation Tasks. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Macau, China, 3–8 November 2019; pp. 6015–6021. [CrossRef]
32. Allshire, A.; Martín-Martín, R.; Lin, C.; Manuel, S.; Savarese, S.; Garg, A. Laser: Learning a Latent Action Space for Efficient Reinforcement Learning. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Xi'an, China, 30 May–5 June 2021; pp. 6650–6656. [CrossRef]
33. Zhang, Q.; Zhu, L.; Chen, Y.; Jiang, S. Constrained DRL for Energy Efficiency Optimization in RSMA-Based Integrated Satellite Terrestrial Network. *Sensors* **2023**, *23*, 7859. [CrossRef] [PubMed]
34. Wang, G.; Wu, F.; Zhang, X.; Guo, N.; Zheng, Z. Adaptive Trajectory-Constrained Exploration Strategy for Deep Reinforcement Learning. *Knowl.-Based Syst.* **2024**, *285*, 111334. [CrossRef]
35. Li, A.C.; Chen, Z.; Klassen, T.Q.; Vaezipoor, P.; Toro Icarte, R.; McIlraith, S.A. Reward Machines for Deep RL in Noisy and Uncertain Environments. *arXiv* **2024**, arXiv:2406.00120v3. [CrossRef]

36. Lillicrap, T.; Hunt, J.; Pritzel, A.; Heess, N.; Erez, T.; Tassa, Y.; Silver, D.; Wierstra, D. Continuous Control with Deep Reinforcement Learning. *arXiv* **2015**, arXiv:1509.02971.
37. Sumiea, E.H.; Abdulkadir, S.J.; Alhussian, H.S.; Al-Selwi, S.M.; Alqushaibi, A.; Ragab, M.G.; Fati, S.M. Deep Deterministic Policy Gradient Algorithm: A Systematic Review. *Heliyon* **2024**, *10*, e30697. [CrossRef] [PubMed]