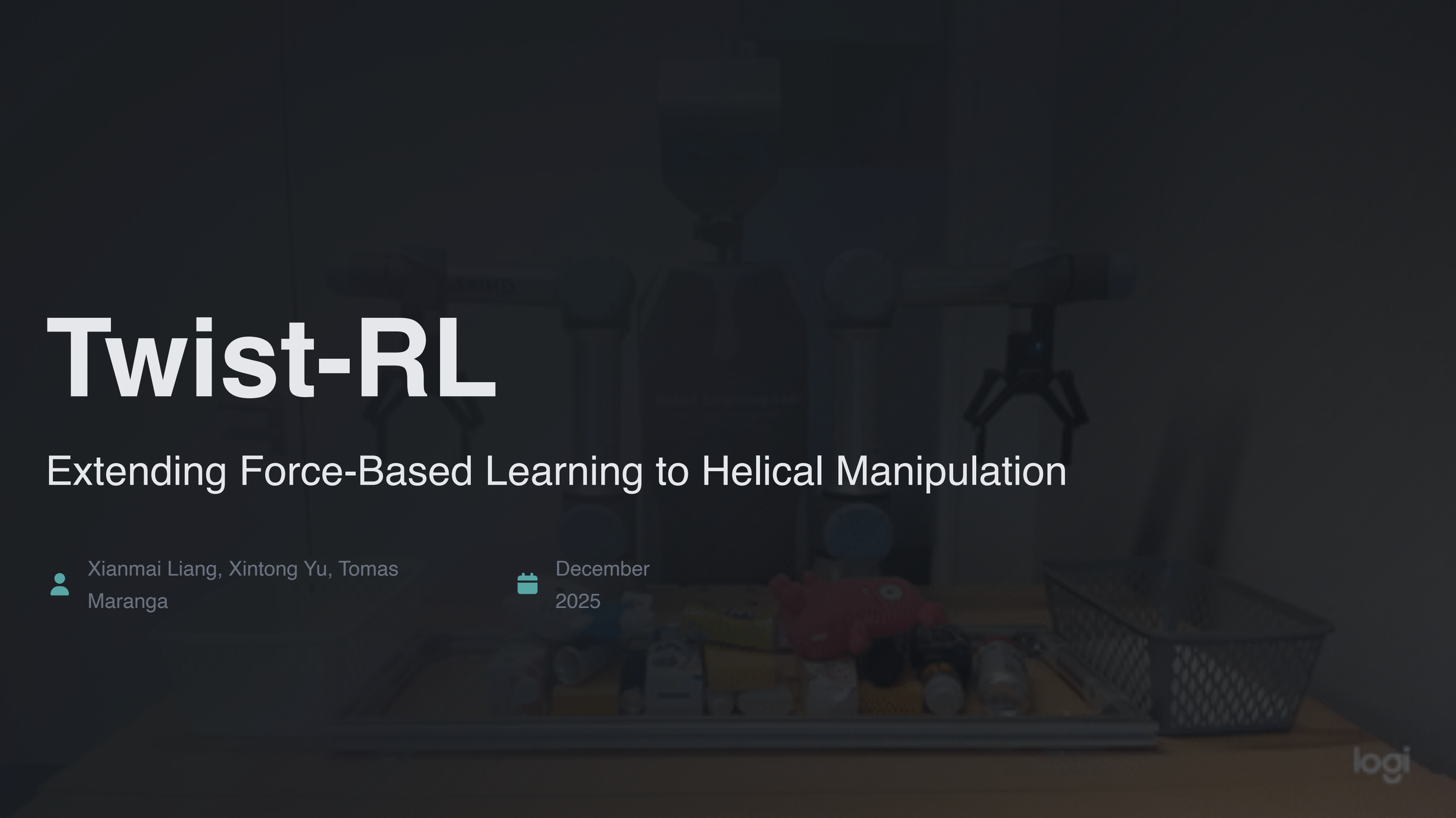
Twist-RL

Extending Force-Based Learning to Helical Manipulation

Xianmai Liang, Xintong Yu, Tomas Maranga

December 2025



Research Overview

01
Introduction
Helical manipulation challenge and research motivation

02
Related Work
FLEX framework, real-world RL, AO-Grasp, and unfastening research

03
Technical Background
MuJoCo/Robosuite platform and TD3 algorithm details

04
Experimental Design
Assumptions, setup, and MDP formulation for helical motion

05
Results & Analysis
Learning performance, critic loss, and Q-value dynamics

06
Future Work
Extensions and similar applications

INTRODUCTION

The Helical Manipulation Challenge

Research Question

Can we extend force-based learning ideas to helical/screw motion?

Proposed Approach

Address by **learning the force space of the object**—formulate object dynamics into transition function and state space, then apply learned force space to similar objects.

This robot-agnostic approach uses force and state at each timestep to define actions, instead of defining joint configurations.

Real-World Application

Approach accounts for real-world deployment by **distributing RL learning into asynchronous processes**, enabling practical application across various robotic platforms.

Motion Types Comparison

Prismatic

Drawers, sliding doors

Revolute

Doors, hinges

Helical/Screw

Bottle caps, screws

RELATED WORK

FLEX Framework for Force-Based Manipulation

FLEX Overview

FLEX: A Framework for Learning Robot-Agnostic Force-based Skills introduces RL-based approach for acquiring force-based manipulation skills that generalize across robotic platforms.

Models object dynamics with prismatic and revolute joints, leveraging Robosuite simulation to learn force representations for each object state.

Key Innovation

Formulates manipulation as **RL problem in force space**—robot autonomously infers joint configurations and interaction dynamics without robot-specific kinematic assumptions. Enables efficient learning and strong generalization.

Our Extension

Inspired by FLEX, we simplified the screw problem by simply considering and learning the torque of our example object. We may define the screwing process using one-degree complexity.

FLEX Framework Components

1. **Object Dynamics Modeling**
   Prismatic & revolute joints
2. **Reinforcement Learning**
   In force space
3. **Robot-Agnostic Policy**
   Generalizes across platforms

Supported Joints

- **Prismatic**
- **Revolute**

Fang, Shijie, et al. *FLEX: A Framework for Learning Robot-Agnostic Force-Based Skills Involving Sustained Contact Object Manipulation*. 19 May 2025, pp. 4782–4788, https://doi.org/10.1109/icra55743.2025.11127866. Accessed 20 Dec. 2025.

Related Work

Real-World RL Integration

Setting up a Reinforcement Learning Task with a Real-World Robot presents framework for integrating RL with physical robotic systems.

Formulates real-world reacher task where robotic arm is trained to reach target positions in 3D space using RL.

Architecture

Separation between **environment thread** and **agent thread** communicating through actuator module. Reduces latency and improves stability.

AO-Grasp Framework

AO-Grasp: Articulated Object Grasp Generation proposes learning-based framework for generating stable, actionable grasps on articulated objects.

Reasons over 6-DoF using segmented partial point clouds, focusing on articulated components while filtering background geometry.

Training

Supervised learning on **~78,000 articulated objects**. Grasp success defined by actionability detection, enabling generalization.

Unfastening Screws

Addresses automated fastener removal in remanufacturing with **robust control strategy for unscrewing** under uncertainty.

Employs spiral search strategy to locate and engage screws despite positional/orientational errors.

Mahmood, A Rupam, et al. “Setting up a Reinforcement Learning Task with a Real-World Robot.” 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 1 Oct. 2018, pp. 4635–4640, https://doi.org/10.1109/iros.2018.8593894. Accessed 11 Apr. 2025.
Morlans, Carlota Parés, et al. AO-Grasp: Articulated Object Grasp Generation. 14 Oct. 2024, pp. 13096–13103, https://doi.org/10.1109/iros58592.2024.10802558. Accessed 20 Dec. 2025.
Li, Ruiya, et al. “Unfastening of Hexagonal Headed Screws by a Collaborative Robot.” IEEE Transactions on Automation Science and Engineering, 2020, pp. 1–14, ieeexplore.ieee.org/abstract/document/8954893, https://doi.org/10.1109/TASE.2019.2958712. Accessed 3 Mar. 2020.

TECHNICAL BACKGROUND

Simulation Platform: MuJoCo & Robosuite

FLEX MDP Formulation

FLEX frames sustained-contact manipulation as **two MDPs**:

Mp = ⟨Sp, A, R, Tp, γ⟩
Prismatic joints

Mr = ⟨Sr, A, R, Tr, γ⟩
Revolute joints

- **State:** Joint axis as 3D unit vector + task progress
- **Action:** Bounded force vector
- **Transition:** Implicit via MuJoCo physics

Key Advantage

Robot-agnostic formulation through object and force-centric approach.

Transition functions Tp and Tr are simply calls to physics engine.

Helical Joint Strategy

MuJoCo suggests **composing hinge joint + slide joint**, then coupling with joint equality:

This effectively yields a **screw joint**. Our project follows this strategy, building helical joint in simulation and training policy on fixed simulator transition.

Deployment Pipeline

1. Train in MuJoCo
2. Deploy to Robosuite
3. Real-world deployment

TECHNICAL BACKGROUND

TD3 Algorithm: Twin Delayed Deep Deterministic Policy Gradient

Core Components

TD3 is an actor-critic RL algorithm for continuous control learning two functions:

Actor $\pi_\phi(s)$
Maps states to actions

Critic $Q_\theta(s, a)$
Estimates expected return

- **Critic training:** Satisfies Bellman equation $Q(s,a) = r + \gamma Q(s', a')$
- **Actor training:** Maximizes critic estimate via policy gradient

TD3's Three Key Improvements

1. **Clipped Double Q-Learning**
   Uses **two critics**, takes minimum $y = r + \gamma \min(Q_1, Q_2)$ to combat overestimation bias
2. **Delayed Policy Updates**
   Updates actor **less frequently** than critics (every $d = 2$ steps), allowing Q-estimates to stabilize
3. **Target Policy Smoothing**
   Adds **clipped noise** to target actions $\tilde{a}' = \pi(s') + \text{clip}(\varepsilon, -c, c)$, preventing Q-function peak exploitation

Training Mechanisms

- **Replay buffer B:** Stores transitions (s,a,r,s',done), samples mini-batches
- **Target networks θ':** Provide stable targets, updated via Polyak averaging

Polyak averaging: $\tau \approx 0.005$

Why TD3 for Twist-RL?

- ✓ Continuous control for torque commands
- ✓ Stable learning for contact-rich tasks
- ✓ Robust value estimation with twin critics

EXPERIMENTAL DESIGN

Assumptions & Setup

1 Grasped Start

Robot's end-effector is **already in contact with and stably grasping** the bottle cap/screw. We don't learn grasping or approach

2 Post-Breakaway

Pre-breakaway. We encoded friction at the start of the unscrewing process to simulated the initial, tightly fastened portion.

3 Known Screw Axis

Assume we can express screw axis in world/object frame as **3D unit vector h** $\in$ $\mathbf{R}^3$. In our case, axis is vertical: [0, 0, 1].

4 Sim-First Approach

All training in MuJoCo/Robosuite with helical joint created as hinge + slide with equality constraint.

5 Mass Considerations

Don't account for item's mass, therefore **don't account for linear/angular acceleration**. Focus is on learning wrench policy.

Problem Focus

These assumptions keep problem focused on **learning wrench policy** for helical manipulation.

✓ Grasping	✓ Approach
✓ Friction	✓ Mass

EXPERIMENTAL DESIGN

MDP Formulation for Helical Motion

$$\text{MDP } M_h = \langle S_t, A, R, T, \gamma \rangle$$

After studying screw theory, we model task as MDP specialized for helical motion.

State Space S_t

$$s_t = [h, p]$$

$h \in \mathbb{R}^3$: Unit vector for screw axis (vertical: [0,0,1])

$p \in \mathbb{R}^3$: Relative contact point position

Action Space A

Axial torque instead of 3D force:

$$a_t = \tau_z$$

Bounded: $\tau_z \in [-\tau_{max}, \tau_{max}]$, allowing scaling to [0,1]

Transition T

Do not learn T. Environment advances via MuJoCo—helical constraint turns applied wrench into coupled rotation/translation.

Returns next state (θ_{t+1}, z_{t+1}).

Reward Function R

Combination of **three** reward types:

1. **Success Reward**
   Large reward after successful unscrew
2. **Progress Reward**
   Based on relative position
3. **Negative Force Reward**
   Minimize torque

Discount γ

$$\gamma = 0.99$$

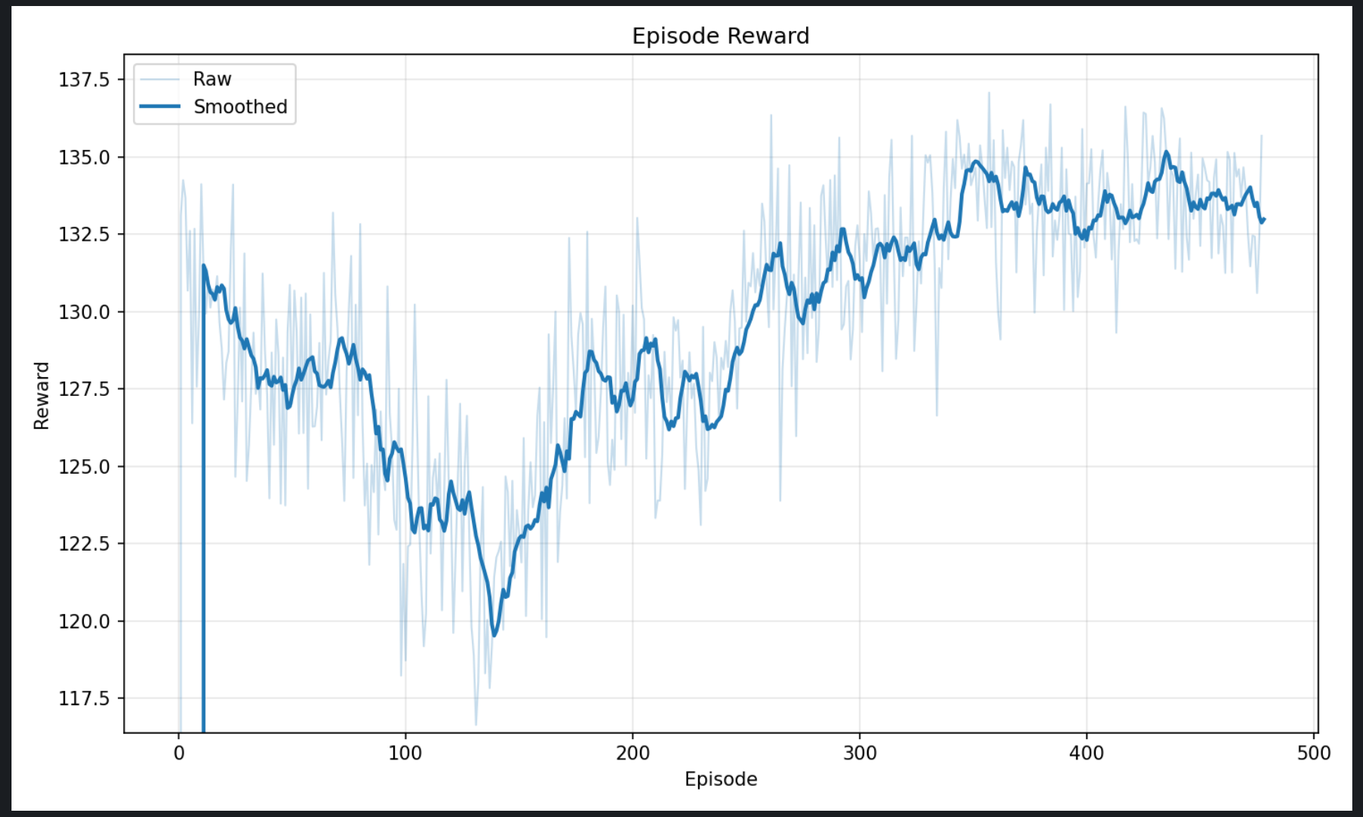
Values **long-term rewards** almost as much as short-term. Success takes sustained effort.

EXPERIMENTAL RESULTS

Learning Performance & Policy Convergence

Episode Reward Over Training

Critic Loss



Training Dynamics Overview

Exploration (0-50 Agent quickly learns effective policy, reaching ~132 reward by episode 50

Significant performance collapse occurs between episodes 100-175, dropping to ~116

Full recovery and stabilization around 131-133 reward after episode 200

Performance Dip Analysis

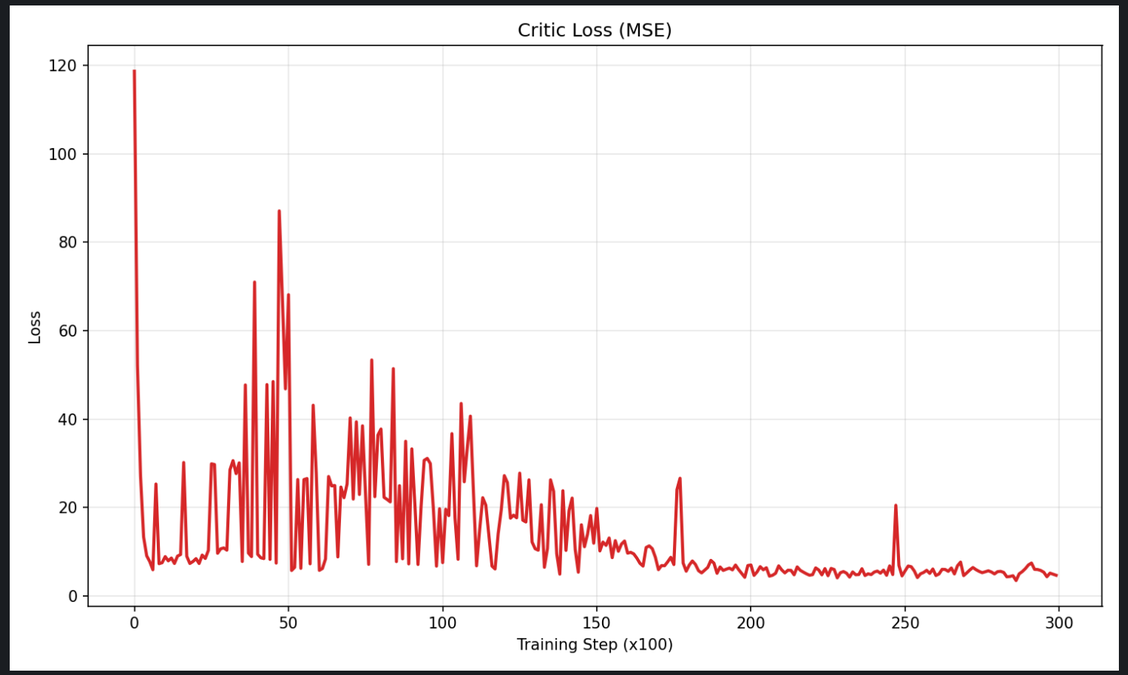
- Likely caused by exploration of suboptimal policy regions
- May indicate temporary instability in the learning process
- Persistent variance in raw rewards reflects inherent environment stochasticity

EXPERIMENTAL RESULTS

Critic Loss & Q-Value Dynamics

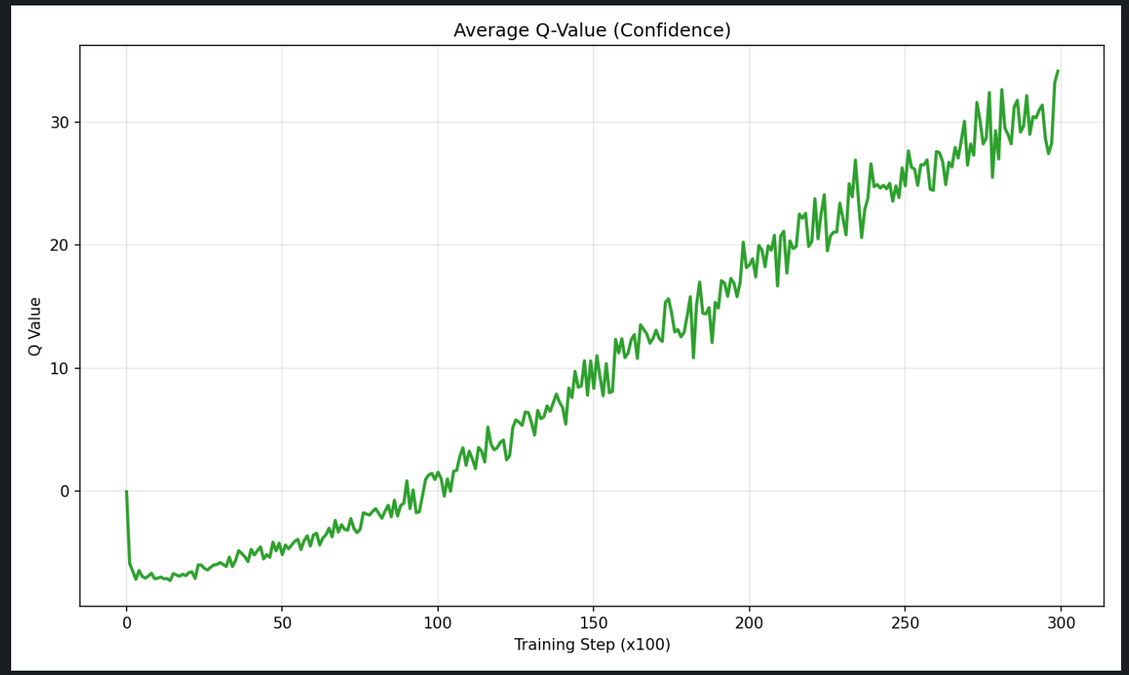
Critic MSE Loss

Critic Loss



Average Q-Value

Avg Q-Value



Early Training

Large loss spikes reflecting instability as critic adapts to changing policy and sparse rewards.

Later Training

Magnitude and frequency decrease as policy stabilizes. Loss remains bounded.

Q-Value Behavior

Monotonic growth indicates increasing confidence but suggests potential overestimation.

DISCUSSION

Key Insights & Implications

Overall Assessment

TD3 rapidly learns stable policy for twisting task, achieving reliable performance after relatively few training episodes.

The framework successfully extends force-based learning to helical manipulation, demonstrating practical applicability.

Key Findings

1. **Reward Plateau**
   Convergence to local optimum, below target
2. **Increasing Q-Values**
   Divergence suggests residual optimism
3. **Rapid Convergence**
   Efficient behavior discovery (~50 episodes)

Implications

Need for Additional Incentives

Observed plateau suggests **policy represents locally optimal solution**. Additional incentives required for continued improvement.

Reward Design Importance

Findings highlight **critical importance of reward design** and exploration strategies for continuous-control RL in contact-rich tasks.

Future Directions

Explore **better exploration strategies**, improved reward shaping, and multi-objective optimization for globally optimal solutions.

FUTURE WORK

Extensions & Applications

Current State

Trained **object-agnostic policy** fitting any object with helicoidal joint. With limited knowledge and perception, model efficiently predicts unknown parameters.

Currently works in Robosuite simulation with ranka Emika Panda arm.

Future Extensions

- Test on other simulation models
- Deploy on real-world robots and objects
- Incorporate friction modeling
- Account for object mass and acceleration
- Optimize for varying grasp positions
- Handle different handle shapes

CONCLUSION

Research Contributions & Impact

Key Contributions

1

Helical Motion Extension

First extension of **force-based learning paradigm to helical motion**.

2

Torque Modeling

Successfully modeled **one-dimensional torque** required for screw-joint interactions.

3

Rapid Policy Convergence

Demonstrated **TD3's efficiency** in learning stable policies for contact-rich manipulation.

Practical Applicability

Robot-Agnostic Design

Approach generalizes across **various robotic platforms** without requiring robot-specific kinematic assumptions.

Real-World Deployment

Future work in asynchronous processes would enable **practical real-world application** with reduced latency and improved stability.

Object Generalization

Learned force space applies to **similar objects with helicoidal joints**.

THANK YOU

Questions & Discussion

Twist-RL: Extending Force-Based Learning to Helical Manipulation

Xianmai Liang, Xintong Yu, Tomas Maranga

December 2025