In-Hand Manipulation with Enforced Grasp Stability for Contact-Rich Tasks

Yifei Chen*, Shihan Lu*, Haoxuan Zhang, and Kevin M. Lynch

***Abstract*—Recent advances in dexterous in-hand manipulation have demonstrated impressive object reorientation capabilities, yet often rely on palm support or fingertip lifting without achieving truly stable grasps. Inspired by the role of rich fingertip force-torque sensing in human manipulation, we propose to integrate force-torque measurements into the observation space of reinforcement learning (RL) policies for robotic in-hand manipulation. To further guide the policy in utilizing contact information, we introduce a force-closure-based reward that explicitly encourages grasp stability during training. We validate our approach on a simulated cube reorientation task using a downward-facing multi-fingered robotic hand, comparing policies trained with and without force/torque observations and grasp stability rewards. Preliminary results suggest that incorporating force-torque sensing and force-closure evaluation improves grasp stability, facilitates smoother in-hand rotations, and accelerates early training progress. These findings highlight the potential of embedding physical grasp constraints into policy learning, though further experiments are needed to fully assess robustness and generalization.**

***Index Terms*—Dexterous manipulation, grasp stability, force closure.**

I. Introduction

In-hand dexterous manipulation has made impressive progress in recent years, enabling robotic hands to reorient objects such as cubes with high levels of flexibility and skill [1]–[3]. However, a common characteristic shared by many of these demonstrations is the reliance on either palm support or fingertip lifting to stabilize the object during manipulation [4], [5]. While effective for controlled environments, these setups often lack a truly stable grasp, meaning the object is not securely enclosed by fingers but instead precariously balanced. As a result, when subjected to unexpected external contacts or disturbances, the grasp can easily fail, causing the object to slip or fall.

In contrast, when humans perform in-hand manipulation, they routinely maintain a stable grasp even while external forces act on the object [6]. A key enabler of this robustness is the continuous perception of contacts exerted on the object through the force and torque sensing at the fingertips, allowing adaptive finger-gait adjustments to maintain secure contact [7]. Compared to other sensing modalities such as visual and tactile feedback (*i.e.* vision-based tactile sensing, or called tactile images [8]), force-torque information between the object and robotic hand's fingertips offers compact, interpretable measurements that correlate strongly with grasp stability, making it an ideal representation for maintaining robust grasps.

Despite the importance of fingertip force-torque sensing in human in-hand manipulation, existing robotic manipulation studies, specially those through Reinforcement-Learning (RL) for finger-gait control, have rarely leveraged this biological insight. Instead, prior work on model-free RL methods for complex in-hand manipulation tasks has predominantly relied on vision, tactile images, or indirect joint-torque measurements as input observations when operating upward or downward-facing multi-fingered hands, leaving the grasping vulnerable to dynamic uncertainty and external contacts [9], [10]. For example, using visual observation, Chen *et al.* [9] utilized teacher-student distillation to transfer policies trained with privileged simulation data to policies in the real world. Sievers *et al.* [10] explored purely tactile-based policies using joint torque feedback in the robotic hand controller, showing that joint-level contact information can enable surprisingly robust manipulation without any vision. However, none of these works have systematically investigated how fingertip-level force-torque observations can contribute to grasp stability in dynamic, contact-rich environments.

We propose to directly integrate force-torque information at robotic fingertips into the observation space of the RL policies for in-hand manipulation. To explicitly encourage grasp stability during the policy training with newly added force-torque data at fingertips, we also introduce a force-closure evaluation metric as part of the reward function, guiding the policy to not only accomplish the task but also maintain a mechanically stable grasp throughout the interaction. Force closure, a fundamental concept in grasp analysis, refers to a grasp configuration where the fingers can resist arbitrary external wrenches through internal forces [11]. Achieving force closure ensures that the object remains securely grasped, irrespective of perturbations, and is essential for real-world robust manipulation. Our ultimate goal is to enhance the grasp stability of the learned policies, enabling them to resist external wrenches while performing dexterous reorientation tasks, such as twisting on/off a jar or using a screwdriver.

Our work aims to bridge this gap by embedding physical grasp constraints into the RL pipeline for in-hand manipulation. By leveraging force-torque feedback and force-closure evaluation, we seek to train policies that achieve more stable, reliable, and physically grounded manipulation behaviors, laying the foundation for future capabilities where robots can perform dexterous in-hand tasks under external disturbances. As an initial effort toward this goal, we focus on in-hand object

*First two authors contributed equally.

All authors are affiliated with Center for Robotics and Biosystems at Northwestern University.

TABLE I
OBSERVATION SPACE COMPOSITION

Name	Notation
Joint angles	$q \in \mathbb{R}^n$
Joint errors	$e_q = q_{\text{cmd}} - q$
Force/torque sensor values	$\mathbf{F}_{\text{tip}} \in \mathbb{R}^{6 \times N}$
Last action (previous command)	$a_{\text{prev}} \in \mathbb{R}^n$
Cube pose and orientation	$p_{\text{cube}} \in \mathbb{R}^3$, $R_{\text{cube}} \in \text{SO(3)}$
Cube linear and angular velocities	$(v_{\text{cube}}, \omega_{\text{cube}}) \in \mathbb{R}^6$
Contact positions and contact forces	$(p_{\text{contact}}, f_{\text{contact}})$

reorientation using a downward-facing multi-fingered robotic hand as our testing platform, where the robotic hand has to address the contact uncertainty introduced by object's gravity.

II. METHOD

A. Task and Environment Setup

The objective is to enable a palm-down robotic hand to grasp a cube and rotate it around its Z-axis as much as possible without dropping it. The task is implemented in the MuJoCo simulator, using a rigid multi-fingered robotic hand (Allegro Hand V4) equipped with 6D force/torque sensors embedded between the fingertip and the last link. Prior work suggests that the initial state of each episode is important for agent to get start with the learning [9], [10], so we deliberately set the initial state of the scene with a relative close grasp and set starting period with zero gravity to allow the agent to have time to react.

B. Observation and Action Spaces

1) Observations: In the Mujoco simulation, we obtain rich information, including the robot states, object (cube) states, and contact force/torque between the object and fingertips. The policy receives all states above as observations during training, as shown in Table I.

To improve temporal coherence and stability, the observations are stacked over 5 frames before being fed into the policy network.

2) Actions: The action space consists of desired joint positions $q_{\text{target}} \in \mathbb{R}^n$, where $n = 16$. Actions are executed through a joint-space PID controller, where the policy outputs target joint positions, and the PID controller handles low-level torque control to track these targets.

C. Reward Design

The overall reward is composed of a *basic task reward* and a *force-closure related grasp stability reward*. The basic task reward includes:

1) Rotation reward: We apply a positive reward proportional to the cube's incremental rotation around the Z-axis between steps.
2) Drop penalty: A large negative reward (-100) is applied if the cube is dropped, and the episode is terminated immediately.
3) Translation and other rotation penalties: Punishments are applied if the cube undergoes significant translations or rotations around axes other than the Z-axis, to encourage controlled spinning.

The base task reward is formulated as follows:

$$\begin{aligned} R_{\text{basic}} = \lambda_{rotate}\Delta\theta - \lambda_{\text{drop}}\text{drop} - \lambda_{\text{z}}|\Delta z| \\ - \lambda_{\text{plane}}|\Delta xy| - \lambda_{\text{rot}}(|\Delta\phi| + |\Delta\psi|) \end{aligned} \tag{1}$$

where each λ is reward weight coefficient corresponding to a specific component.

In contrast to palm-supported manipulation, a single mistake can easily break grasp closure and result in object dropping. Exploration must balance risk (lifting fingers for regrasping) and grasp stability, making the training prone to local minima without careful design.

To enhance grasp stability during manipulation, we introduce force-closure related reward to encourage the agent to learn from the fingertip-level force/torque information:

1) Contact reward: We apply a positive reward proportional to the number of fingers in contact with the object. This encourages maintaining multiple active contacts to increase stability.
2) Force-closure (FC) metric reward: We compute a grasp quality score based on the grasp wrench matrix G, constructed from the measured contacts, as the following two steps:

First, based on the contact information, we build the candidate wrench matrix. For each contact, we generate a set of candidate contact forces inside the discretized friction cone. For each candidate force, we compute the associated wrench (force and torque relative to the object's center of mass). All candidate wrenches are aggregated into a matrix $G \in \mathbb{R}^{6 \times N}$. Second, we evaluate the grasp quality by performing SVD on G to obtain singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_6$. The smallest singular value σ_{min} reflects the minimum resisted wrench magnitude, which is later normalized into a FC score $\in [0, 1]$ [11], [12].

The final form of the overall reward is:

$$R = R_{\text{basic}} + \lambda_{\text{contact}} \times \text{contact count} + \lambda_{fc} \times \text{FC score} \tag{2}$$

where λ_{contact} and λ_{fc} are weights of corresponding contact-related rewards.

D. RL Pipeline

We use the Soft Actor-Critic (SAC) algorithm [13] for policy learning. The network uses a fully connected neural networks for both actor and critic, with two hidden layers of 512 neurons each. We use 12 parallel simulation environments to accelerate data collection. The training was conducted on a laptop equipped with an Intel Core i7 CPU, an NVIDIA RTX 3060 Laptop GPU. Each policy was trained for approximately 2 million environment steps, requiring about 8 hours of wall-clock time.

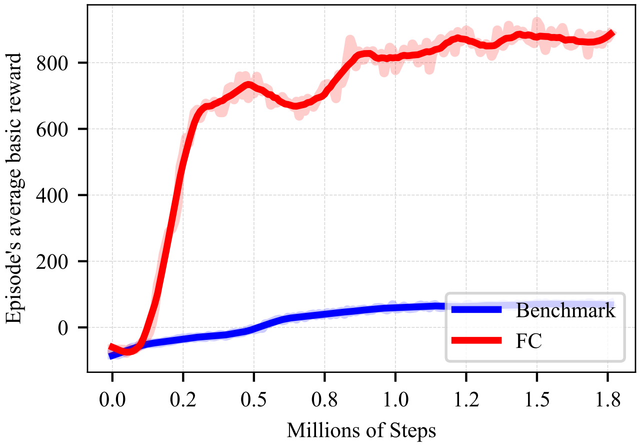

Fig. 1. Average episode basic task reward versus training steps (in millions) for benchmark policy and force-closure (FC) enhanced policy with additional force/torque observations. The FC policy quickly outperforms the benchmark in terms of average basic task reward, surpassing 600 reward by 0.5 millions steps and converging around 850 reward. The benchmark slowly reaches under 100 reward over the entire training horizon. The shaded envelopes around the curves denote the variability across runs.

III. Results

A. Experimental Setup

We aim to evaluate the impact of force/torque sensing and force-closure rewards on in-hand manipulation stability while pursuing object orientation. In the preliminary experiment, three settings are compared:

1) **Baseline:** No force/torque sensing; basic task reward.
2) **Force/Torque Only:** Force/torque observations added; basic task reward.
3) **Force/Torque + Force-Closure:** Force/torque observations added, along with additional force-closure related reward.

All policies are trained under identical conditions using SAC. Evaluation focuses on two key metrics: (1) Episode length—the time before object drop—indicating grasp stability, and (2) Accumulated rotation—the cumulative Z-axis rotation achieved during an episode, representing in-hand dexterity.

Policies are trained without environmental randomization. During evaluation, we perform rollouts both with and without initial state perturbations. Specifically, we introduce noise by randomly perturbing the initial cube position (±10 mm) and orientation (±5°). Each policy is evaluated over 50 episodes, with a maximum episode length of 10 seconds.

B. Experimental Results

As shown in Fig. 1, the episode's average basic task reward curves during training indicated that introducing force/torque sensing and force-closure rewards improves episode duration of the training compared to the baseline policy. It enabled the agent to maintain more stable contact with the cube across the episode and provides greater flexibility for exploration.

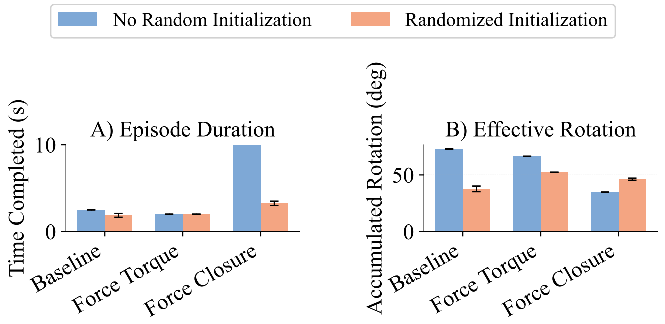

Fig. 2. Evaluation results comparing different policies under both nominal and randomized initializations. (A) Episode duration—the time until the cube drops—demonstrates stability. (B) Effective accumulated rotation—measured only when at least two fingers maintain contact with the cube—reflects in-hand manipulation ability. Policies trained with force-closure rewards exhibited better trade-off between stability and dexterity.

Additionally, policies incorporating FC rewards demonstrated faster initial learning progress compared to the baseline policy, achieving significant reward improvements at early stage of the training. This suggested that embedding physical grasp constraints into the learning process may accelerate policy optimization toward more stable manipulation strategies.

During policy rollout (Fig. 2), when evaluating without random initialization, the proposed FC policy showed significantly extended episode duration with a decreased effective rotation angle, partially due to the dramatic spinning motions observed in the baseline policy or policy without FC reward. After adding initial state perturbations, FC policy continued to achieve longer episode durations with only a minor reduction on the effective rotation, compared to the policies without the specific FC reward. However, it still achieved greater effective rotation than the baseline policy.

Moreover, as the time lapse illustrated in Fig. 3, policies trained with force-closure rewards exhibited more consistent in-hand rotation behavior and fewer grasp failures, compared to baseline policies that tended to spin the cube in the air, merely pursuing larger angular changes but ignoring the grasp closure.

It is important to note that these findings are based on preliminary experiments. While promising trends are observed, further evaluations, including additional randomization during training and tests under external disturbances, are needed to fully validate the generalization and robustness benefits of the proposed method.

IV. Discussion

Our preliminary experimental results suggest that incorporating force/torque information into the observation space along with physics-informed contact rewards helps the agent to better balance grasp stability and object reorientation objectives. While promising trends are observed, further experiments and adjustments are still needed to reach solid conclusions.

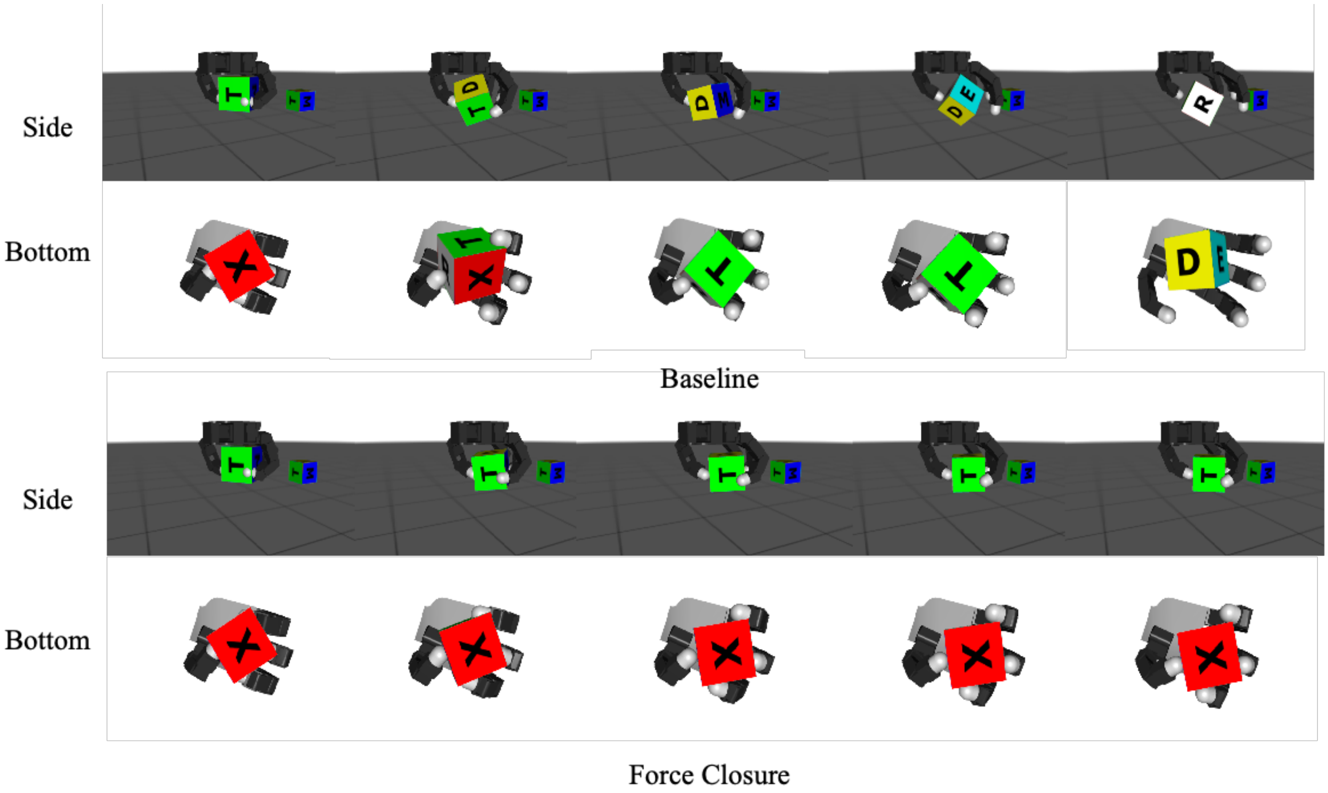

Fig. 3. Time-lapse comparison of manipulation behavior for different policies. Each row represents a different policy, while columns correspond to snapshots at increasing time steps (400 ms apart). Policies trained with force closure rewards demonstrate significantly better grasp stability while rotating the cube. In contrast, the baseline policy tends to spin the object in the air, leading to unstable grasps and object drops.

While evaluating the results, we observe that the baseline policy used for comparison does not achieve ideal performance. Our baseline reward function and environment setup were designed to closely follow prior work from the TUM AIDX Lab series [10], [14], [15]. However, differences in the dexterous hand model, low-level joint controller, and initial learning conditions likely contributed to deviations from the originally reported results. This discrepancy suggests that further reward tuning, controller calibration, and initialization adjustments are necessary to establish a stronger benchmark for fair comparison. It also reveals the sensitivity of the trained policies in the absence of enforced grasp stability constraints.

Looking forward, future work will focus on applying external disturbances to better evaluate the agent's grasp stability under challenging conditions. We also plan to investigate the sim-to-real transfer performance of the proposed approach to assess its practical applicability in the real world.

References

[1] OpenAI, I. Akkaya, M. Andrychowicz, M. Chociej, M. Litwin, B. McGrew, A. Petron, A. Paino, M. Plappert, G. Powell, R. Ribas, J. Schneider, N. Tezak, J. Tworek, P. Welinder, L. Weng, Q. Yuan, W. Zaremba, and L. Zhang, "Solving rubik's cube with a robot hand," 2019. [Online]. Available: https://arxiv.org/abs/1910.07113

[2] Z.-H. Yin, C. Wang, L. Pineda, F. Hogan, K. Bodduluri, A. Sharma, P. Lancaster, I. Prasad, M. Kalakrishnan, J. Malik, M. Lambeta, T. Wu, P. Abbeel, and M. Mukadam, "Dexteritygen: Foundation controller for unprecedented dexterity," 2025. [Online]. Available: https://arxiv.org/abs/2502.04307

[3] H. Qi, B. Yi, M. Lambeta, Y. Ma, R. Calandra, and J. Malik, "From simple to complex skills: The case of in-hand object reorientation," *arXiv preprint arXiv:2501.05439*, 2025.

[4] G. Khandate, M. Haas-Heger, and M. Ciocarlie, "On the feasibility of learning finger-gaiting in-hand manipulation with intrinsic sensing," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2022, pp. 2752–2758.

[5] J. Wang, Y. Yuan, H. Che, H. Qi, Y. Ma, J. Malik, and X. Wang, "Lessons from learning to spin" pens"," *arXiv preprint arXiv:2407.18902*, 2024.

[6] J. R. Flanagan, M. C. Bowman, and R. S. Johansson, "Control strategies in object manipulation tasks," *Current opinion in neurobiology*, vol. 16, no. 6, pp. 650–659, 2006.

[7] D. A. Rosenbaum, J. D. Slotta, J. Vaughan, and R. Plamondon, "Optimal movement selection," *Psychological Science*, vol. 2, no. 2, pp. 86–91, 1991.

[8] E. Donlon *et al.*, "Gelslim: A high-resolution, compact, robust, and calibrated tactile-sensing finger," in *IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2018, pp. 1927–1934.

[9] T. Chen, J. Xu, and P. Agrawal, "A system for general in-hand object re-orientation," in *Conference on Robot Learning*. PMLR, 2022, pp. 297–307.

[10] L. Sievers, J. Pitz, and B. Bäuml, "Learning purely tactile in-hand manipulation with a torque-controlled hand," in *International Conference on Robotics and Automation (ICRA)*, 2022, pp. 2745–2751.

[11] K. M. Lynch and F. C. Park, *Grasping and Manipulation*. Cambridge University Press, 2017, p. 400–444.

[12] M. A. Roa and R. Suárez, "Grasp quality measures: Review and performance," in *Journal of Autonomous Robots*, 2015, pp. 65–88.

[13] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," in *International conference on machine learning*. Pmlr, 2018, pp. 1861–1870.

[14] J. Pitz, L. Röstel, L. Sievers, and B. Bäuml, “Dextrous tactile in-hand manipulation using a modular reinforcement learning architecture,” in *IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 1852–1858.

[15] L. Röstel, J. Pitz, L. Sievers, and B. Bäuml, “Estimator-coupled reinforcement learning for robust purely tactile in-hand manipulation,” in *IEEE International Conference on Humanoid Robots (Humanoids)*, 2023, pp. 1–8.