Evaluating Dexterity Limitations of Bionic Hands via Reinforcement Learning

Student: ABDEL-WAHAB, Yussef **PK:** 52009226
Peer Reviewer: ABURAIA, Mohamed

Abstract—**In-hand manipulation is a key benchmark for dexterous robotic control, especially in the context of bionic hands, where mechanical simplicity often competes with functional complexity. In this work, Isaac Sim and reinforcement learning (RL) are utilized to evaluate the dexterity of a custom-designed bionic hand by training it to manipulate a cube in 3D space. The objective is to assess whether the current design provides sufficient degrees of freedom (DoF) to complete the task in simulation under controlled conditions with domain randomization. Domain randomization is employed to ensure the potential for zero-shot transfer from simulation to the real world. Despite leveraging a high-fidelity physics engine and robust policy optimization, the bionic hand consistently failed to perform the manipulation task. It is hypothesized that this failure stems not from insufficient learning or simulation inaccuracies, but from limitations in the mechanical design of the hand itself. To validate this, the same simulation environment and training pipeline were tested on a 3D model of the Shadow Robotics Hand, which successfully completed the task. Based on these results and careful observation during training, it was concluded that the current bionic hand model lacks the necessary dexterity to achieve the task. To test this hypothesis, three variants of the bionic hand were evaluated: the baseline configuration, a version with an extended supination range, and a version with an additional wrist flexion/extension joint. Their performance were compared against the 24-DoF Shadow Hand as a reference. The modified bionic hands showed clear improvements over the baseline, isolating the impact of joint range and placement on task success. The results suggest that moderate increases in range of motion along manipulation-relevant axes, or the addition of even a single DoF, can substantially enhance manipulation capabilities, highlighting the importance of joint-level design choices in enabling dexterous bionic and robotic hands.**

Keywords—**Dexterous Robotics, Reinforcement Learning, In-Hand Manipulation, Sim-to-Real**

I. Introduction

Modern robotic systems are typically designed with a specific use case in mind and deployed in controlled, structured environments. While this design has enabled reliable performance in industrial automation and service robotics, it shows clear limitations in human-centric settings, which are characterized by variability, uncertainty, and unstructured interactions [1] [2]. To address this challenge, research has increasingly turned toward humanoid and bionic robotic systems that are capable of operating in environments originally built for humans.

Within this area of research, dexterous robotic hands and arms represent a particularly promising but difficult path. A bionic arm with multiple degrees of freedom can, in principle, replicate human-like manipulation and enable a wide range of tasks. However, the very complexity that makes such systems versatile also creates challenges. High-dimensional control spaces, nonlinear joint couplings, and intricate contact dynamics make classical control methods difficult to apply effectively [2]. As a result, controllability and robustness remain major bottlenecks in the deployment of dexterous robotic systems [1].

Previous attempts to solve the complex nature of high dimensional action spaces required the use of model predictive training or contact-invariant optimization [3] [4]. The inherent issue with in-hand manipulation stems most likely not from the action space alone but also the observation space, including contact points [5].

In recent years, reinforcement learning (RL) has emerged as a compelling alternative for solving these challenges [2], [6]. Unlike traditional controllers, RL agents can learn policies directly from interaction data, making them well-suited for tasks with complex dynamics such as in-hand object manipulation. Landmark studies have demonstrated that RL can control high-degree-of-freedom robotic hands to perform nontrivial manipulation tasks in simulation and, with the help of techniques like domain randomization, even transfer these skills to real-world hardware [7]. These advances suggest that simulation-based RL provides a viable pathway for training dexterous bionic arms to achieve robust, human-like manipulation in diverse environments.

II. State of the Art

Dexterous in-hand manipulation requires coordinated control across many coupled degrees of freedom. In practice, however, many robotic and prosthetic hands provide only a small number of actuated joints. This mismatch forces controllers to achieve outcomes that normally require complex hands with only simplified mechanisms. The following overview highlights how research has approached this problem through contact based planning, learning based control, the exploitation of external forces, and design strategies that rely on either compliance or rigid linkages.

Early approaches to dexterous manipulation treated it as a problem of motion planning with explicit contact reasoning. Contact invariant optimization planned trajectories in combination with contact events, which produced finger gaits and regrasp motions in simulation [4], [5]. Predictive control schemes similarly synthesized rapid behaviors for dynamic manipulation tasks [3]. These methods established as an important foundation, that with accurate models of friction and contact, coordinated motions could be planned even when actuation was limited. However, in real scenarios such approaches were often not sufficient, since minor errors in modeling or friction parameters could destabilize the resulting plans [1], [2]. For hands with few actuators, the dependence on subtle contact exploitation made the gap between simulation and reality especially significant.

This behavior that emerged from these limitations is referred to as extrinsic dexterity. Instead of relying only on the internal joints of the hand, a controller can make use of gravity, inertia, and environmental contact surfaces to reorient an object [8]. Some studies demonstrated how external forces can increase the set of reachable object configurations with minimal finger motion [8]. More recently, learning based methods rediscovered these principles, with policies that pressed or rolled objects against planes or the palm to create rotations that could not be commanded directly [9], [10]. This work makes clear how a simple hand can still perform meaningful reorientation by creating contact situations that substitute for missing degrees of freedom. Parameterized manipulation primitives now provide a structured way to represent such strategies [11].

The rise of deep reinforcement learning introduced a powerful alternative to analytic control. With appropriate reward design and exploration, reinforcement learning agents discovered non intuitive solutions for in-hand reorientation tasks [2]. A landmark example is the OpenAI Shadow Hand project, in which a policy optimized with Proximal Policy Optimization (PPO) [6] successfully reoriented a cube and later solved the Rubik's Cube [7]. The success depended on extensive domain randomization, which randomized masses, frictions, and joint gains to improve robustness [12]. Later work extended this pipeline to more agile behaviors on the same platform, showing that realistic simulation combined with massive randomization can enable direct transfer to hardware [13]. Theoretical analyses now provide guarantees that domain randomization can narrow the gap between simulated and real environments under broad conditions, and that memory dependent policies can further stabilize performance when dynamics vary [14].

Although many of these advances were demonstrated on hands with a large number of actuated joints, the same principles can be applied to simpler devices. Tendon driven designs with rolling contacts, for example, provide natural compliance that makes it easier for policies to exploit contact geometry [15]. At the opposite end, dynamic two handed throw and catch experiments highlight how momentum and timing can replace internal joints with well orchestrated external effects [16]. Recent studies also demonstrate end to end visual policies for humanoid manipulation tasks, showing that learning directly from camera observations can reduce the mismatch between simulation and sensor inputs [17]. Other work combines reinforcement learning with demonstration led curricula, which improves stability during training without altering the mechanical complexity of the platform [18].

The physical mechanism itself ultimately determines what strategies are possible. Tendon driven fingers allow a form of compliant following that can be approximated by proportional derivative joint controllers, where dependent joints naturally lag behind and adjust smoothly. By contrast, rod coupled or gear coupled mechanisms behave as stiff kinematic loops. These must be modeled with extremely high stiffness values and minimal damping, and backlash needs to be represented explicitly to avoid unrealistic rigid couplings. Studies that successfully transferred from simulation to real Shadow Hands emphasize the importance of aligning simulator contact models, joint limits, and controller gains with the actual mechanism [13].

In summary, low actuator hands can still achieve tasks that normally require complex devices, provided that they exploit extrinsic contacts, are trained under diverse simulation conditions, and are modeled with accurate representations of their mechanical couplings. High actuator hands remain the most capable, but targeted design choices such as compliant tendons or fast global wrist joints can raise the requirements for simpler hands. This understanding motivates the following study, in which a rod coupled bionic hand with limited joints is evaluated against a high degree Shadow Hand baseline. The simulations explicitly account for stiff linkage behavior and backlash, and training is performed with PPO in Isaac Sim with extensive domain randomization following established practice.

III. Materials and Methods

In this work, the Bionic Arm developed in [19] was evaluated in the in-hand cube manipulation task and compared against the Shadow Hand [20]. An altered version of the Bionic Arm was also tested, incorporating an additional axis in the wrist. To improve computational efficiency during simulation, the CAD models of both versions of the Bionic Arm were simplified. Fasteners such as screws, nuts, and bolts were intentionally excluded. The outer geometry, particularly the surfaces that may come into contact with the cube, was preserved to avoid altering interaction behavior. Regions previously occupied by washers were filled to prevent edge gaps that could catch the cube and lead to non-transferable behaviors. Internal cavities not involved in contacts were also filled where appropriate. One simplification that may influence sim-to-real transfer is the exclusion of the metallic linkage rods that couple the distal phalanx as depicted in Fig. 1. This rod can come in contact with the cube, since it is mounted on the inside of each finger.

It was left out because it forms a closed-loop mechanism that Unified Robot Description Format (URDF) does not natively support. URDF is an XML-based description language widely used in robotics to specify a robot's kinematic and dynamic properties, including joints, links, and sensor configurations. However, URDF supports only tree-structured kinematic chains and does not represent closed-loop mechanisms such as the four-bar linkage in the Bionic Hand. As a workaround, URDF provides the `mimic` joint feature. In this approach, one joint is defined as dependent on another, with

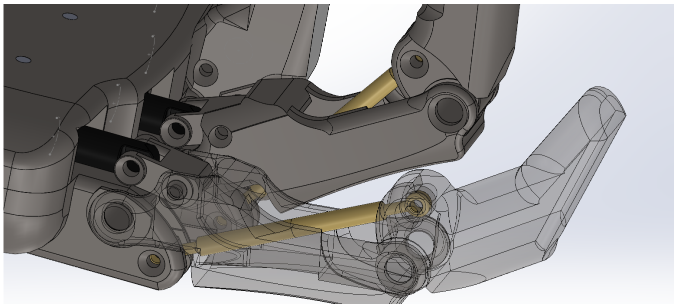


Fig. 1: Steel linkages (yellow) connecting the distant phalanx with the driven middle phalanx.

its position expressed as a linear function of the parent joint's position , as shown in Equation 1.

$$q_{\text{distal}} = \alpha q_{\text{middle}} + \beta, \tag{1}$$

where α is a scaling factor and β an offset. Since four-bar-linkages in robotic fingers constrain the distal joint to a single degree of freedom [21], and the motion of the distal phalanx in the Bionic Arm is a linear function of the middle phalanx, the mimic joint feature can be used as an approximation. In this implementation, the distal phalanx joints were set to mimic the middle phalanx joints with the appropriate scaling factor, effectively reproducing the closed-loop motion in simulation. The assembly can then be exported from SolidWorks using the ROS add-on `sw_urdf_exporter`. The resulting URDF was then imported as a Universal Scene Description (USD) into NVIDIA Isaac Sim for simulation.

Inside Isaac Sim, further properties of the joints and linkages can be configured. Since the connection to the fingertips in the Bionic Hand prototype is established with a stiff metal rod instead of compliant tendons (Fig. 1), the mimic-follow behavior requires a more detailed description. Behind the `mimic` tag, Isaac Sim implements a standard Proportional-Derivative (PD) controller, where the stiffness parameter corresponds to the proportional gain (P-gain) and the damping parameter corresponds to the derivative gain (D-gain) [22]. This formulation is appropriate for tendon-driven linkages, as it allows the dependent joint to give in and lag slightly behind during movement. Since most dexterous robotic hands, such as the Shadow Hand, employ tendon-driven actuation, the PD-based mimic joint is typically sufficient [7].

In contrast, the Bionic Hand prototype uses a rigid rod–linear motor mechanism, where each finger is directly actuated along a single axis. For such stiff kinematics, a conventional PD controller is inadequate, since the mechanism requires an instantaneous response of the driven linkage. To approximate rigid coupling in simulation, the stiffness must be set to a very high (near-infinite) value, while damping is set to zero. This ensures that mimic joints behave as stiff connections, causing the fingertips to follow the driving joint without delay.

However, to enable zero-shot transfer from simulation to the real world, it is also necessary to simulate the backlash introduced by the drive mechanism as accurately as possible. In this case, a Proportional-Threshold (PT) controller would be more appropriate than a PD controller [23] [24]. A PT controller is a type of non-linear control system. Unlike a standard proportional controller that applies a corrective force directly proportional to the position error, a PT controller incorporates a dead zone (threshold). Within this dead zone, the controller output remains zero, meaning no corrective action is applied. Only when the error exceeds this predefined threshold does the controller activate, applying a force proportional to the remaining error [25]. In order to overcome this issue, empirical values for the P- and D-gains were set to approximate the backlash observed on each finger.

A. Reinforcement Learning Configuration

Isaac Sim was used with the PhysX backend to simulate the physics in the environment, while Isaac Lab was employed to manage the RL tasks. For training, the RL library `skrl` [26] was used together with a Proximal Policy Optimization (PPO) agent [6].

The only difference in training between the Shadow Hand and the Bionic Hand was the size and depth of the neural network policy. The Shadow Hand was trained with a four-layer fully connected network of sizes [512, 512, 256, 128], whereas the Bionic Hand employed a deeper five-layer network of sizes [1024, 1024, 512, 256, 128]. The larger architecture was chosen to better handle the increased complexity of the Bionic Hand, which included nonlinearities and backlash effects in its joints, making the policy optimization problem more challenging [27] [28].

The observation space consisted of 120 dimensions, including joint positions and velocities, the cube's position, orientation, and velocity, the goal position and orientation, fingertip positions and velocities, as well as the previous action values. The action space, in contrast, was limited to 7 dimensions, corresponding to the commanded positions of each actuated degree of freedom (DoF). A summary of the observation and action space is provided in Table 1.

Tab. 1: Observation and Action Space Configuration

Space	**Description**
Observations (120D)	Joint positions and velocities; Cube position, orientation, velocity; Goal position and orientation; Fingertip positions and velocities; Previous actions
Actions (7D)	Target positions for each DoF

Each training run consisted of 80,000 simulation steps with 8192 parallel environments. To promote robustness and enable zero-shot transfer from simulation to the real world, domain randomization was applied [12], [29]–[31]. For each environment instance, random values were sampled for the following parameters:

- static and dynamic friction between cube and hand,
- joint stiffness and damping values,
- upper and lower joint limits,

- mass distribution of the cube,
- gravity vector applied to the scene.

The reward function was designed to encourage successful object manipulation. The most important components were:

- **Distance reward:** penalizing deviation of the cube's position from the goal.
- **Rotation reward:** encouraging alignment between cube and goal orientation.
- **Action penalty:** discouraging unnecessarily large or unstable joint actions.
- **Goal bonus:** rewarding successful alignment within tolerance.

Additional shaping terms (e.g., velocity scaling, force/torque scaling, and averaging factors) were tuned empirically to balance stability and exploration. The key reward scaling parameters are summarized in Table 2.

Tab. 2: Reward Function Parameters

Parameter	**Value**
Distance reward scale	-10.0
Rotation reward scale	1.0
Action penalty scale	-0.001
Goal bonus	250
Success tolerance (rad)	0.05
Fall penalty	0
Fall distance	0.24
Velocity obs. scale	0.2
Force/torque obs. scale	10.0
Averaging factor (α)	0.1
Action moving average	0.3

IV. Results

To evaluate the performance of each agent, several parameters were tracked across episodes to compare the different configurations. In particular, the number of successful trials and dropped episodes was recorded for each run. Each episode consisted of 600 steps and was terminated either when the object was dropped or when the maximum step count was reached. Episodes that did not result in a success or a drop were classified as other.

A distinction was made when classifying cases in which both the success and drop conditions were triggered. For the analysis of episode outcomes, any episode in which the cube was successfully reoriented at least once was counted as a success. In contrast, when evaluating reward values or episode length, such cases were labeled as drops.

In addition, the rotational error was computed and monitored throughout each episode, along with the corresponding return values.

To ensure stability of the simulation and avoid introducing artifacts, the training of the Bionic Hand in its standard configuration, as presented in [19], was conducted in stages. The process began with the kinematic model and joint limits, while actuator dynamics were left unconstrained. In the next step, stiffness and damping values were explicitly defined for the joints. Subsequently, joint velocity limits were enforced to better approximate the physical system. Finally, simulations were repeated with backlash enabled on the joints and domain randomization applied to capture additional mechanical characteristics of the real hardware.

With this staged approach, the standard configuration was able to sequentially solve the cube rotation task. However, once actuator speed limits were introduced, the same level of performance was no longer achievable. It was observed that the agent initially relied on rapid, momentum-based reorientation motions, a strategy that became infeasible once actuator velocity limits were enforced.

Training the 7-DoF Bionic Hand with all limitations and domain randomization enabled proved unsuccessful without further modifications. The system appeared to lack the ability to maneuver the cube into more favorable positions within the workspace of each finger, which prevented complex reorientations. As a result, the agent became trapped in a reward plateau. Instead of attempting reorientation, which often led to dropping the cube, it learned to balance the cube on the palm to maximize reward.

To address this, the wrist rotation was extended to include an additional $90°$ of supination beyond the typical human anatomical range, resulting in a full $270°$ range of motion. This modification was motivated by observations from training the 8-DoF version of the Bionic Hand that also included dorsiflexion and palmarflexion of the wrist. In that case, although the added DoFs enabled more varied strategies, the limited supination range still caused difficulties: depending on the randomized initial pose and velocity, the cube often rolled toward the hypothenar region, from which recovery was impossible. Consequently, the agent adopted an alternative strategy of balancing the cube on the back of the thumb, where the available $90°$ of supination could be exploited to stabilize, recover and even reorient the cube.

A. Bionic Hand (7-DoF)

The Bionic Hand was also evaluated under full physical constraints, including domain randomization, joint backlash, and velocity limits. Under these conditions, the task could not be solved. In rare cases, when the randomized initial state was favorable, the hand was able to balance the cube on its palm. However, complex reorientation maneuvers were not attempted, as the limited actuation strategies provided few options for moving the cube into configurations that enabled further manipulation.

As illustrated in Figure 2, the system converged to a steady state after approximately 80,000 training timesteps, with no further manipulations attempted. Based on these results, it can be concluded that the initialization of the cube led to frequent early drops before the hand could react. This effect was reinforced by the velocity limits and by initializing the robotic hand in a random configuration. Without the additional supination range, the palm's orientation was often misaligned with the cube's initial dropping velocity vector, causing deflections on impact and leading to premature episode termination.

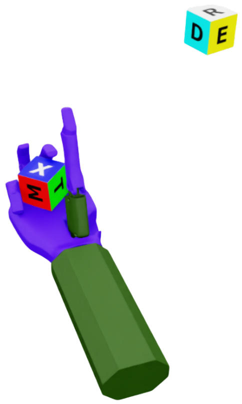


Fig. 2: Steady state of the 7-DoF Bionic Hand after 80,000 timesteps of training, where the cube remains balanced on the palm without further manipulation.

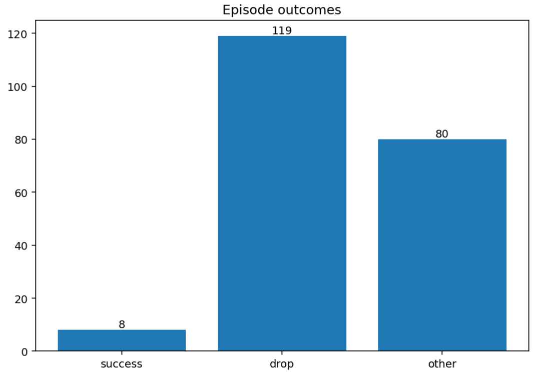


Fig. 3: Outcome distribution for 200 evaluation episodes of the 7-DoF Bionic Hand under full constraints.

Figure 3 shows that in 80 out of 200 episodes the cube remained balanced on the palm. Of the 200 episodes, 8 were registered as successful completions. However, 7 of these 8 cases were immediately followed by a drop, indicating that they were likely favorable initializations rather than genuine solves. This interpretation is supported by the time-to-success distributions in Figure 4, where most successes occurred unrealistically early. Only one episode reached completion at around 100 steps, which aligns more closely with the mean time-to-success observed in the extended supination configuration.

Overall, these results demonstrate that without the additional supination range the Bionic Hand was unable to reliably achieve stable reorientation.

B. Bionic Hand (7-DoF with Extended Supination)

The reorientation strategy of the 7-DoF Bionic Hand with Extended Supination relied strongly on extrinsic dexterity [8]–[10]. Whenever possible, the cube was allowed to rest on the palm while the distal phalanx of the thumb was positioned against the proximal phalanx of either the middle or index finger, creating a confined space that prevented the cube from slipping (Figure 9). Within this configuration, wrist flicks combined with coordinated finger positioning enabled clockwise or counterclockwise rotations of the cube. For flipping maneuvers, the thumb in conjunction with the index and middle fingers was used to push the cube against the metacarpal region of the thumb, after which a simultaneous wrist flick and thumb swivel produced the rotation. Once a coarse orientation was achieved, the hand attempted fine alignment by forming a firmer grasp [32]. The configuration illustrated in Figure 9a represents the firmest grasp achievable by the Bionic Hand, where the cube is lifted from the palm and stabilized in the target orientation.

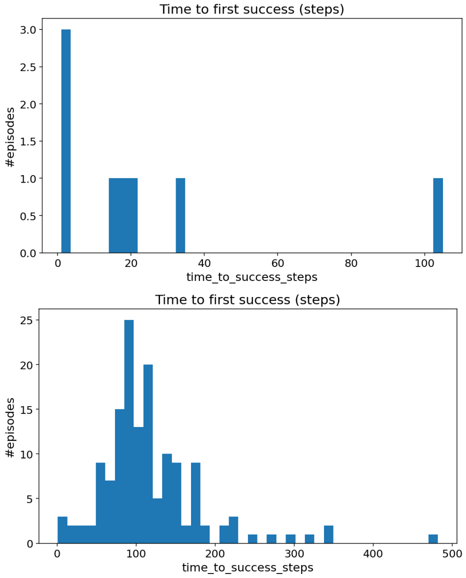


Fig. 4: Comparison of time-to-success distributions. Top: 7-DoF Bionic Hand under full constraints. Bottom: 7-DoF Bionic Hand with extended supination.

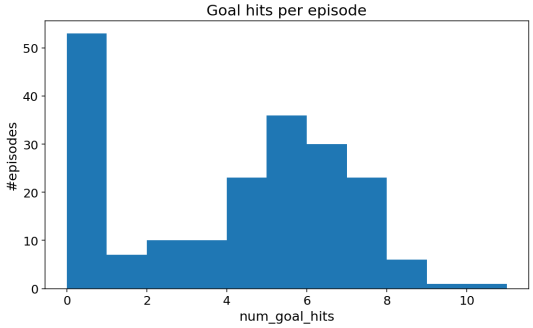


Fig. 5: Distribution of goal hits per episode for the 7-DoF Bionic Hand with extended supination.

In contrast, the Shadow Hand relied less on palm support and more on direct finger coordination. Its higher number of degrees of freedom enabled stable in-hand reorientation without the need to exploit extrinsic forces to the same extent, allowing for more consistent manipulation strategies [7] [1].

Figure 5 shows the distribution of goal hits per episode during evaluation. While a large number of episodes achieved zero goal hits, the majority of successful episodes clustered around 4 to 8 hits, with a peak at 5 to 6. This indicates that the agent, when successful, was able to repeatedly reorient the cube within the same episode.

The relationship between return and minimum rotation error per episode is illustrated in Figure 6. Successful episodes (blue) dominate the region close to the defined success tolerance (22.9°). Dropped episodes (orange) are associated with high rotation errors and consistently low returns, meaning that fewer drops occur after the task has been solved once. This confirms that the return function strongly correlates with orientation accuracy, and that the additional supination increased the likelihood of stabilizing the cube within the tolerance region.

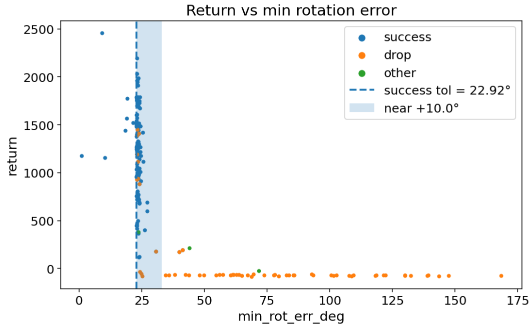


Fig. 6: Return as a function of minimum rotation error per episode.

Finally, Figure 7 shows the distribution of minimum rotation errors across episodes. Most successful runs achieved errors very close to the success threshold, with a dense concentration between 20° and 25°. A smaller fraction of episodes resulted in much higher errors ($> 40°$), typically corresponding to dropped cubes. The shaded region in the plot highlights the tolerance and near-miss zones, confirming that the agent was able to reliably achieve cube orientations near the success threshold, though occasional instability remained.

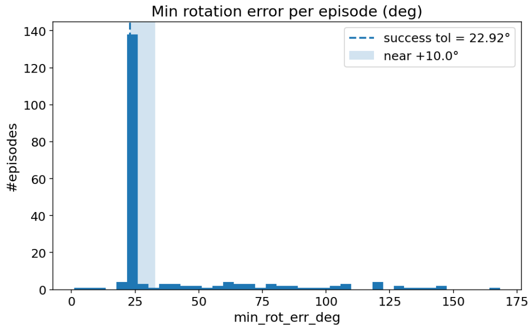


Fig. 7: Distribution of minimum rotation error per episode.

Overall, the introduction of an additional 90° of supination improved the agent's ability to stabilize the cube and avoid manipulation dead-ends. The results suggest that increased wrist mobility is a critical factor for achieving reliable in-hand manipulation, as it reduces the dependence on momentum-driven reorientation strategies.

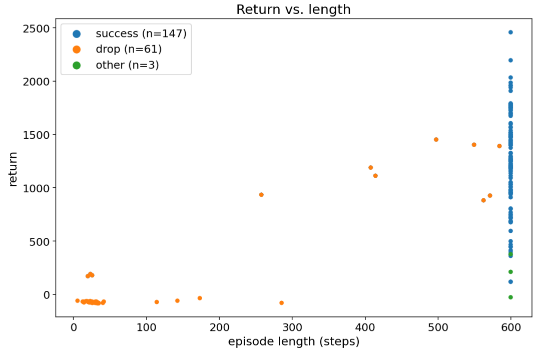
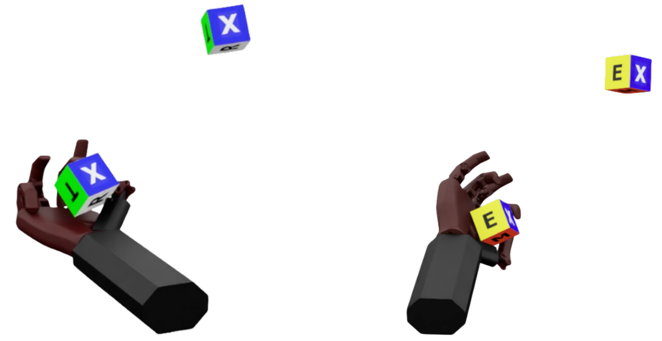


Fig. 8: Return as a function of episode length. Instances where the cube converged close to the target orientation but failed to complete the reorientation task are highlighted in green.

As illustrated in Figure 8, the majority of cube drops occurred early in the episodes, suggesting that the initialization of cube states could be further optimized for the Bionic Hand. Additionally, several episodes showed instances where the agent achieved an initial success but subsequently dropped the cube before completion. Across all evaluated episodes, 147 were classified as successful, 61 as drops, and 3 as near-misses.



(a) 7-DoF with extended supination Bionic Hand failing to reach the goal

(b) 7-DoF with extended supination Bionic Hand failing to reach the goal due to domain randomisation.

Fig. 9: Performance of the 7-DoF extended supination Bionic Hand in two scenarios.

As shown in Figure 9, the agent occasionally came very close to fully solving the cube reorientation task. However, in the example on the right (Figure 9b), the full 90° supination range had already been utilized, constraining the agent in a configuration from which further reorientation was not possible. Careful inspection reveals that the achieved range does not perfectly align with the intended 90° supination. This discrepancy arises from domain randomization: during training, joint limits were randomized across environments, and the agent learned behaviors under slightly different constraints. When evaluated in this instance, the learned strategy, effective in an environment with looser limited joint paramters, led the agent to attempt a rotation that exceeded the randomized limit of the current environment. As a result, the policy was trapped at the boundary of the joint's feasible range, unable to complete the reorientation.

C. Bionic Hand (8-DoF with Wrist Flexion/Extension)

Even in this configuration, without enabling the extension of the supination, the results were still lacking. Since this version was designed as an exaggerated configuration of the Bionic Hand, the extra 90° of supination was also unlocked. With this additional freedom, the agent focused on solving the task in the most efficient way possible. The added axis allowing for flexion/extension of the wrist, together with the extended supination, provided relatively fast actuation speeds (180°/s for flexion/extension and 230°/s for supination, compared to only 80°/s for the fingers). As a result, the policy relied mainly on wrist motions to orient the cube, while the fingers were mainly used for stabilization and small corrective adjustments.

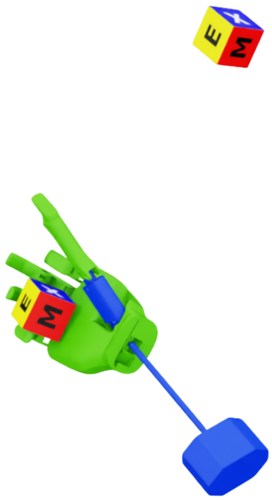


Fig. 10: 8-DoF Bionic Hand with wrist flexion/extension using the fast axes to orient the cube and keeping it in place with the finger joints.

The evaluation outcomes are summarized in Figure 11. Out of all episodes, 183 were classified as successful, while 33 resulted in drops.

The distribution of minimum rotation errors per episode is shown in Figure 12. Most successful trials achieved errors tightly clustered around the 22.9° tolerance, with a sharp peak between 20° and 25°. A smaller fraction of episodes produced much higher errors (> 40°), which typically corresponded to drops.

Figure 13 shows the time-to-first-success distribution. Most successful episodes converged within 50–100 steps, indicating that the additional wrist DoFs enabled rapid alignment strategies. Only a small number of cases required more than 200 steps, typically when the cube was initialized in an unfavorable pose.

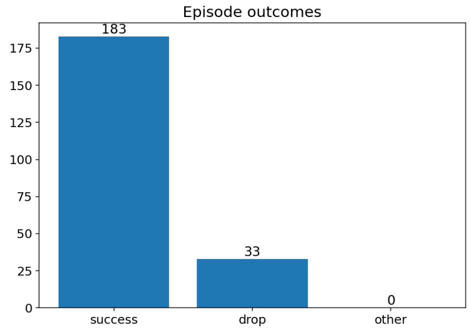


Fig. 11: Outcome distribution for the 8-DoF Bionic Hand with wrist flexion/extension.

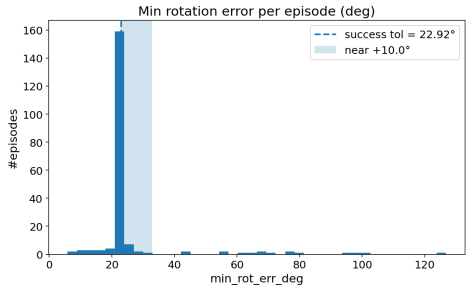


Fig. 12: Distribution of minimum rotation errors for the 8-DoF Bionic Hand.

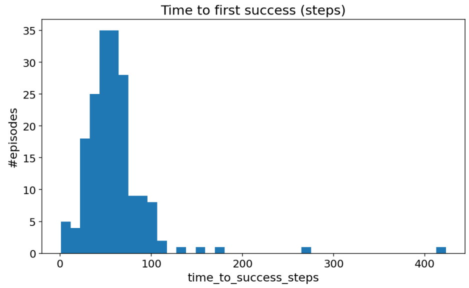


Fig. 13: Time to first success in the 8-DoF Bionic Hand.

Finally, Figure 14 shows return as a function of episode length. Most successful episodes cluster near the maximum length, reflecting policies that achieved repeated reorientations. In contrast, drops typically occurred earlier and yielded lower returns, whereas later drops were associated with higher returns, indicating that the agent often lost the cube only after several successful maneuvers.

Overall, the 8-DoF configuration demonstrated a clear improvement in stability and efficiency compared to the 7-DoF versions. The fast wrist axes dominated the manipulation strategy, enabling rapid coarse reorientations, while the fingers provided fine control.

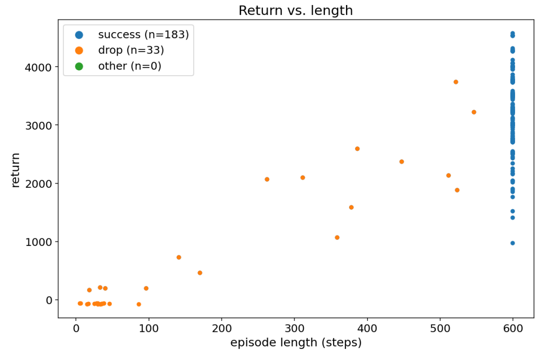


Fig. 14: Return versus episode length for the 8-DoF Bionic Hand.

D. *Shadow Hand (24-DoF)*

Figure 15 presents the distribution of goal hits per episode during evaluation. In contrast to the Bionic Hand, the Shadow Hand achieved a broader spread of goal hits, with a significant portion of episodes exceeding ten successful reorientations. This suggests that the higher number of degrees of freedom allow for more frequent and consistent in-hand reorientation cycles within a single episode.

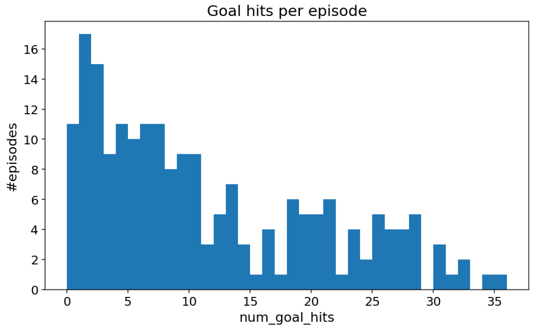


Fig. 15: Distribution of goal hits per episode for the Shadow Hand.

The distribution of minimum rotation errors (Figure 16) further supports this observation. Most successful episodes achieved errors tightly clustered near the success threshold, with relatively few cases exceeding 40°. This concentration highlights the Shadow Hand's ability to consistently bring the cube into near-goal orientations without relying on boundary strategies.

Finally, Figure 17 illustrates the relationship between return and episode length. Most episodes terminated with either a clear success or a failure, suggesting that the Shadow Hand policy rarely exhibited unstable or oscillatory behavior. Successful episodes (blue) form a dense cluster around the defined success tolerance (22.9°), whereas drops (orange) are more evenly distributed across the return range. In comparison to the Bionic Hand, the distribution of dropped cases is more dispersed, with only a small fraction associated with low returns. Drops occurring later in the episode correspond to higher returns, indicating that the agent often completed several successful reorientations before ultimately losing the cube. This behavior was observed only when using the modified `inhand_manipulation_env.py` in combination with the adapted `play.py` script that enabled explicit tracking of successes. In contrast, when evaluated with the original environment configuration used for training and the un-edited `play.py`, drops were virtually absent but could not be systematically recorded.

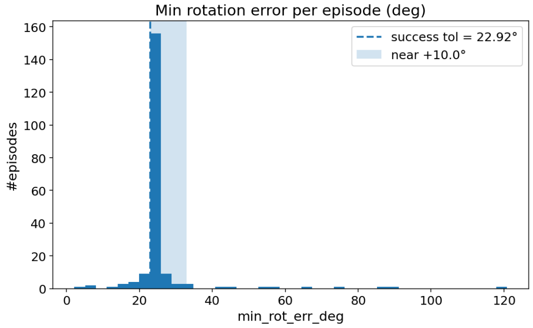


Fig. 16: Distribution of minimum rotation error per episode for the Shadow Hand.

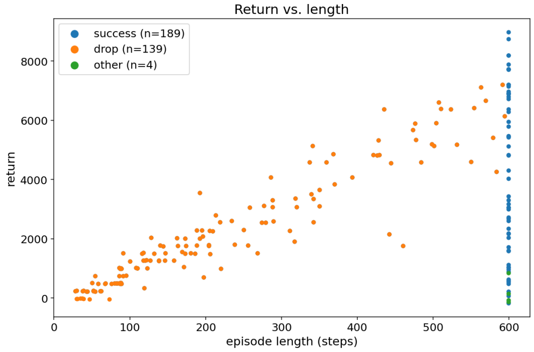


Fig. 17: Return as a function of episode length for the Shadow Hand. Episodes that converged to kinematic deadlock states are highlighted in green.

Overall, the Shadow Hand achieved more frequent and stable reorientation cycles than the Bionic Hand, with tighter rotation error distributions and fewer near-miss outcomes. Some dropped episodes emerged, however these typically followed several successful reorientations. The results highlight that higher degrees of freedom increase robustness in cube stabilization but also expand the policy search space, making training outcomes more sensitive to reward design and exploration.

V. Discussion

The results indicate that the primary bottleneck is mechanical rather than algorithmic. The 7-DoF baseline plateaued under realistic actuator limits, favoring balance-and-hold over true reorientation. The findings suggest that wrist mobility plays a central role in enabling stable in-hand manipulation while minimizing reliance on high-velocity, impulse-based motions. As expected for lower intrinsic dexterity, the bionic variants leaned on extrinsic contacts (palm and thumb–finger corridor), while the Shadow Hand required less environmental support.

Simulation fidelity limits remain: linear `mimic` with very high P and near-zero D suppresses backlash, some initializations left little corrective margin and broad domain-randomization ranges can teach strategies that overfit to joint-limit "slack." Although a deeper network was used for the bionic hand to account for nonlinearities, training both hands with the same architecture would rule out network size as a factor in the performance gap.

Practically, modest and well-placed DoF/range changes can deliver large gains. Fast orthogonal wrist axes plus finger stabilization proved especially effective. Further steps inlcude modelling the four-bar as parallel kinematics with a dead-zone (backlash) controller in USD, switch from simulator pose to vision estimates for sim-to-real, diversify objects/contacts, and report stability-oriented metrics (time-in-tolerance, multi-success streaks, actuation costs), culminating in zero-shot hardware validation.

VI. Summary and Outlook

In this work, a custom bionic hand was evaluated to determine whether it can be trained to perform dexterous in-hand cube manipulation using reinforcement learning. A reproducible pipeline was established in Isaac Lab: CAD models were exported to URDF (with linear `mimic` couplings), converted to USD, and used to train PPO agents via `skrl`. Since URDF cannot represent closed kinematic loops, the distal–middle phalanx coupling was approximated with linear mimic joints. Within Isaac Sim, this induces a PD controller with stiffness and damping mapped to P/D gains. For stiff rod-driven fingers, near-infinite stiffness and minimal damping were required to emulate rigid linkages; however, this configuration removes backlash. As a result, a Proportional–Threshold (PT) controller with a dead-zone was identified as a promising alternative for future work to better capture transmission free-play.

From a learning perspective, the 7-DoF baseline could balance the cube but failed to consistently reorient it once actuator speed limits were imposed. In the absence of actuator constraints, the policies exploited high-velocity, momentum-transfer strategies, but performance plateaued once these were restricted. Extending wrist supination by 90° (to 270° total) enabled recovery postures and repeatable reorientation cycles, improving success rates and reducing kinematic deadlock situations. Adding wrist flexion/extension (resulting in an 8-DoF hand) further enhanced stability and efficiency. For comparison, the Shadow Hand (24-DoF) achieved the tightest rotation-error distributions and frequent multi-success episodes, though it occasionally exhibited late drops under the evaluation protocol that explicitly logged successes.

Future work can improve both the modeling and evaluation of the bionic hand. On the modeling side, the four-bar linkage with a stiff rod could be represented as parallel kinematics directly in USD, thereby capturing contact interactions more faithfully. A custom backlash-controller may also be integrated to more accurately reproduce the true mechanical free-play.

On the learning side, training could be extended with convolutional neural networks that estimate the cube pose from virtual camera input, rather than relying on ground-truth data from Isaac Sim, to facilitate sim-to-real transfer. Beyond single-object tasks, future experiments could incorporate multiple objects, varied masses, and different friction properties. Reporting standardized metrics—such as success rate, median rotation error, time-to-success, multi-success streaks, and actuation costs (energy/torque)—would further strengthen comparability across studies. Finally, validation on real hardware with zero-shot transfer would provide the ultimate benchmark for the proposed approach.

Overall, the results show that extending supination significantly improved success and recovery, while adding flexion/extension increased efficiency. The 7-DoF hand relied heavily on external forces such as gravity to achieve reorientation, whereas the Shadow Hand required less environmental support.

References

[1] G. Li, R. Wang, P. Xu, Q. Ye, and J. Chen, "The developments and challenges towards dexterous and embodied robotic manipulation: A survey," 2025. [Online]. Available: https://arxiv.org/abs/2507.11840

[2] C. Yu and P. Wang, "Dexterous manipulation for multi-fingered robotic hands with reinforcement learning: A review," *Frontiers in Neurorobotics*, vol. Volume 16 - 2022, 2022. [Online]. Available: https://www.frontiersin.org/journals/neurorobotics/articles/10.3389/fnbot.2022.861825

[3] V. Kumar, Y. Tassa, T. Erez, and E. Todorov, "Real-time behaviour synthesis for dynamic hand-manipulation," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, 2014, pp. 6808–6815.

[4] I. Mordatch, Z. Popović, and E. Todorov, "Contact-invariant optimization for hand manipulation," in *ACM SIGGRAPH/Eurographics Symposium on Computer Animation*. Eurographics Association, 2012, pp. 137–144.

[5] I. Mordatch, E. Todorov, and Z. Popović, "Discovery of complex behaviors through contact-invariant optimization," *ACM Trans. Graph.*, vol. 31, no. 4, Jul. 2012. [Online]. Available: https://doi.org/10.1145/2185520.2185539

[6] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *CoRR*, vol. abs/1707.06347, 2017. [Online]. Available: http://arxiv.org/abs/1707.06347

[7] OpenAI, M. Andrychowicz, B. Baker, M. Chociej, R. Jozefowicz, B. McGrew, J. Pachocki, A. Petron, M. Plappert, G. Powell, A. Ray, J. Schneider, S. Sidor, J. Tobin, P. Welinder, L. Weng, and W. Zaremba, "Learning dexterous in-hand manipulation," 2019. [Online]. Available: https://arxiv.org/abs/1808.00177

[8] N. C. Dafle, A. Rodriguez, R. Paolini, B. Tang, S. S. Srinivasa, M. Erdmann, M. T. Mason, I. Lundberg, H. Staab, and T. Fuhlbrigge, "Extrinsic dexterity: In-hand manipulation with external forces," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, 2014, pp. 1578–1585.

[9] W. Zhou and D. Held, "Learning to grasp the ungraspable with emergent extrinsic dexterity," 2022. [Online]. Available: https://arxiv.org/abs/2211.01500

[10] C. Ma, H. Yang, H. Zhang, Z. Liu, C. Zhao, J. Tang, X. Lan, and N. Zheng, "Dexdiff: Towards extrinsic dexterity manipulation of ungraspable objects in unrestricted environments," 2024. [Online]. Available: https://arxiv.org/abs/2409.05493

[11] S.-M. Yang, M. Magnusson, J. A. Stork, and T. Stoyanov, "Learning extrinsic dexterity with parameterized manipulation primitives," 2024.

[12] J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, and P. Abbeel, "Domain randomization for transferring deep neural networks from simulation to the real world," 2017. [Online]. Available: https://arxiv.org/abs/1703.06907

[13] A. Handa, A. Allshire, V. Makoviychuk, A. Petrenko, R. Singh, J. Liu, D. Makoviichuk, K. Van Wyk, A. Zhurkevich, B. Sundaralingam, Y. Narang, J.-F. Lafleche, D. Fox, and G. State, "Dextreme: Transfer of agile in-hand manipulation from simulation to reality," 2024.

[14] X. Chen, J. Hu, C. Jin, L. Li, and L. Wang, "Understanding domain randomization for sim-to-real transfer," 2022, publisher Copyright: © 2022 ICLR 2022 - 10th International Conference on Learning Representationss. All rights reserved.; 10th International Conference on Learning Representations, ICLR 2022 ; Conference date: 25-04-2022 Through 29-04-2022.

[15] Y. Toshimitsu, B. Forrai, B. G. Cangan, U. Steger, M. Knecht, S. Weirich, and R. K. Katzschmann, "Getting the ball rolling: Learning a dexterous policy for a biomimetic tendon-driven hand with rolling contact joints," in *Humanoids*, 2023.

[16] B. Huang, Y. Chen, T. Wang, Y. Qin, Y. Yang, N. Atanasov, and X. Wang, "Dynamic handover: Throw and catch with bimanual hands," 2023.

[17] T. Lin, K. Sachdev, L. an, J. Malik, and Y. Zhu, "Sim-to-real reinforcement learning for vision-based dexterous manipulation on humanoids," 2025, project: https://toruowo.github.io/recipe.

[18] M. Bauza, J. E. Chen, V. Dalibard, N. Gileadi, R. Hafner, M. F. Martins, J. Moore, R. Pevceviciute, A. Laurens, D. Rao, M. Zambelli, M. Riedmiller, J. Scholz, K. Bousmalis, F. Nori, and N. Heess, "Demostart: Demonstration-led auto-curriculum applied to sim-to-real with multi-fingered robots," 2024.

[19] F. Dannereder and P. H. Pachschwoll, "Development of a 3d-printed bionic hand with muscle- and force control," in *Austrian Robotics Workshop 2018*, 2018, pp. 59–66.

[20] Shadow Robot Company, "Shadow Robot Dexterous Hand," https://www.shadowrobot.com/products/dexterous-hand/, 2005, accessed: [18.09.2025].

[21] M. West, "Four-bar linkages - dynamics," https://dynref.engr.illinois.edu/aml.html, 2015, accessed on 26.06.2025.

[22] NVIDIA. (2025) Tuning gains. NVIDIA Corporation. [Online]. Available: https://docs.isaacsim.omniverse.nvidia.com/latest/robot_setup/ext_isaacsim_robot_setup_gain_tuner.html

[23] R. Bruns, J. Diepstraten, X. Schuurbiers, and J. Wouters, *Motion control of systems with backlash*, ser. DCT rapporten. Technische Universiteit Eindhoven, 2006, dCT 2006.075.

[24] F. Cursi, W. Bai, E. Yeatman, and P. Kormushev, "Model learning with backlash compensation for a tendon-driven surgical robot," *IEEE Robotics and Automation Letters*, vol. 7, pp. 1–8, 07 2022.

[25] J. Vörös, "Modeling and identification of systems with backlash," *Automatica*, vol. 46, no. 2, pp. 369–374, 2010.

[26] A. Serrano-MuÃ±oz, D. Chrysostomou, S. BÃ¸gh, and N. Arana-Arexolaleiba, "skrl: Modular and flexible library for reinforcement learning," *Journal of Machine Learning Research*, vol. 24, no. 254, pp. 1–9, 2023. [Online]. Available: http://jmlr.org/papers/v24/23-0112.html

[27] S. Levine, C. Finn, T. Darrell, and P. Abbeel, "End-to-end training of deep visuomotor policies," 2016. [Online]. Available: https://arxiv.org/abs/1504.00702

[28] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015. [Online]. Available: https://doi.org/10.1038/nature14236

[29] T. Dai, K. Arulkumaran, T. Gerbert, S. Tukra, F. Behbahani, and A. A. Bharath, "Analysing deep reinforcement learning agents trained with domain randomisation," *Neurocomputing*, vol. 493, pp. 143–165, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231222003708

[30] M. Ma, H. Li, G. Hu, S. Liu, and D. Zhao, "Comparison of different domain randomization methods for policy transfer in reinforcement learning," in *2023 IEEE 12th Data Driven Control and Learning Systems Conference (DDCLS)*, 2023, pp. 1818–1823.

[31] A. Shakerimov, T. Alizadeh, and H. A. Varol, "Efficient sim-to-real transfer in reinforcement learning through domain randomization and domain adaptation," *IEEE Access*, vol. 11, pp. 136 809–136 824, 2023.

[32] C. Pehoski, A. Henderson, and L. Tickle-Degnen, "In-hand manipulation in young children: rotation of an object in the fingers," *American Journal of Occupational Therapy*, vol. 51, no. 7, pp. 544–552, 1997.