

Composing Dexterous Grasping and In-hand Manipulation via Scoring with a Reinforcement Learning Critic

Lennart Röstel* Dominik Winkelbauer* Johannes Pitz Leon Sievers Berthold Bäuml

Abstract—In-hand manipulation and grasping are fundamental yet often separately addressed tasks in robotics. For deriving in-hand manipulation policies, reinforcement learning has recently shown great success. However, the derived controllers are not yet useful in real-world scenarios because they often require a human operator to place the objects in suitable initial (grasping) states. Finding stable grasps that also promote the desired in-hand manipulation goal is an open problem.

In this work, we propose a method for bridging this gap by leveraging the critic network of a reinforcement learning agent trained for in-hand manipulation to score and select initial grasps. Our experiments show that this method significantly increases the success rate of in-hand manipulation without requiring additional training. We also present an implementation of a full grasp manipulation pipeline on a real-world system, enabling autonomous grasping and reorientation even of unwieldy objects. Website: aidx-lab.org/manipulation/icra25

I. INTRODUCTION

Dexterous grasping and in-hand manipulation with multi-fingered hands are challenging tasks involving multi-contact states and high degrees of freedom. While the state-of-the-art for both problems recently benefited from learning-based methods [1, 2], they are still largely addressed independently: Grasping deals with predicting static hand-object configurations that are robust to external disturbances by optimizing for a suitable grasp metric [1]. Meanwhile, in-hand manipulation is about dynamically reorienting the object towards a desired object configuration [3, 4, 5].

While a stable grasp is often a prerequisite for successful in-hand manipulation, optimally combining grasping and in-hand manipulation for end-to-end performance remains an open problem. In particular, a key question is how to select grasps that best promote downstream in-hand manipulation tasks. Assessing this suitability of a given grasp for subsequent manipulation (sometimes referred to as *manipulability*) is traditionally challenging as it depends on a myriad of factors like the initial object pose, the object's shape and physical properties, the desired object configuration, and the hand's kinematic capabilities.

In this work, we propose a simple but effective solution to this problem that uses the critic network from a reinforcement learning agent after training to select grasps suitable for subsequent in-hand manipulation. To our knowledge, we are the first to address this problem with such a general solution, enabling increased end-to-end performance of a

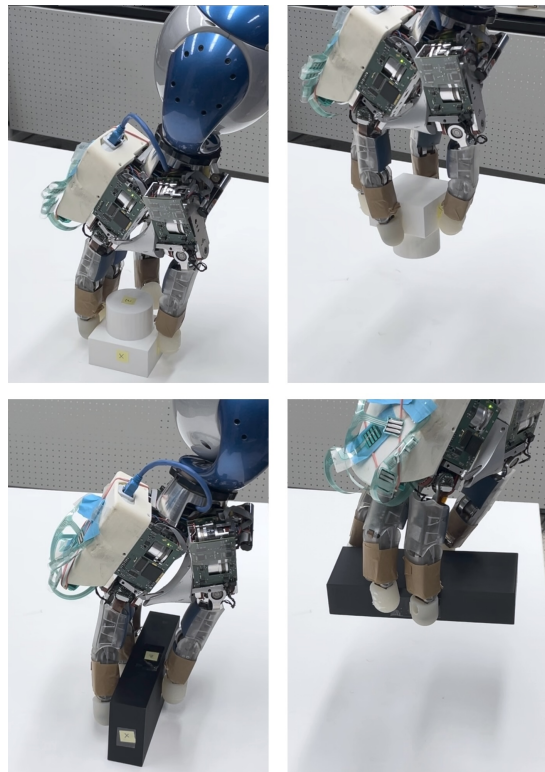


Fig. 1. Grasping and in-hand manipulation of two objects with the DLR Hand II [6]. Left: an object-specific grasp is selected, optimized to facilitate subsequent in-hand manipulation to the desired goal orientation (right).

grasp-manipulation pipeline. Moreover, we think the method described in this paper applies to a larger class of problems in robotics, where optimization over a set of candidate initial states post-training is possible.

A. Related Work

1) *Grasping*: Grasping has been studied extensively in the past. In early works, grasping is approached by analytical analysis using grasping properties like force closure [7] or more fine-grained grasp qualities [8, 9, 10]. With the advent of data-driven approaches, grasps could be found efficiently online based on incomplete observations using SVMs [11], logistic regression [12], or, more recently, deep-learning-based methods [13, 14, 15]. The latter allows high-dimensional input spaces without manual feature engineering and modeling complex grasp distributions for multi-fingered hands [16, 17, 1]. Most mentioned methods are only concerned with the general grasp stability, while our approach further allows selecting grasps based on their suitability for

*Equal contribution. The authors are with the Learning AI for Dexterous Robots Lab, Technical University of Munich, Germany (aidx-lab.org), and the DLR Institute of Robotics & Mechatronics (German Aerospace Center). Contact: {lennart.roestel, dominik.winkelbauer}@tum.de

the consecutive manipulation task. Existing approaches that rate grasps based on their manipulability typically make use of the hand-object jacobian [18]: A manipulability measure can be defined via its condition number [19], minimum singular value [20], or the volume of the spanned manipulability ellipsoid [21]. While the listed methods were designed for serial manipulators, they can be transferred to multi-finger grasps [22]. Later, they were extended to support underactuated hands [23] or to allow more fine-grained requirement specification [24]. While the mentioned approaches allow a more task-specific grasp selection, they only consider the grasp in a quasi-static setting. In contrast, our approach rates grasps based on their suitability for the full subsequent manipulation task, including dynamic effects and finger gaiting.

2) *In-hand manipulation*: OpenAI et al. [3] for the first time successfully applied deep reinforcement learning to reorient objects via finger gaiting. The seminal work showed that learning-based methods could overcome the challenges of modeling fragile multi-contact dynamics and long-horizon planning which classical approaches [25, 26, 27] have struggled with. Since then the field has seen impressive progress with simpler hardware setups, shorter training times [2, 28, 29, 30], for a specific object [31, 32, 5, 30, 33] or for manipulating multiple different objects, including objects unseen during training [34, 35, 36, 37, 38]. In all of these works, objects are placed in the hand by a human operator, limiting their applicability to real-world scenarios. Chen et al. [34] additionally consider the task of manipulating objects that are placed on a table, however, without lifting them up.

In this work, we study in-hand manipulation initialized from diverse stable grasps that can be executed on a real robot system. From an exploration perspective, a diverse distribution of initial states is known to be an important factor for successful reinforcement learning [33]. Additionally, having access to explicitly planned stable grasps (as opposed to learning an end-to-end policy for grasping and manipulation) allows to chain grasping, transportation, and in-hand manipulation in a controlled and generalizable manner.

Existing work on joint optimization of grasping and manipulation is limited to reduced settings like planar manipulators [39]. For the challenging case of multi-fingered hands, in this work for the first time we combine grasping and general in-hand reorientation in 3D in a principled way.

B. Contributions

In summary, we present the following contributions:

- We propose a novel method for scoring dextrous grasps, making use of a reinforcement learning critic network trained for in-hand manipulation.
- Utilizing grasp planning, as opposed to a heuristic across all objects, we can successfully train purely tactile in-hand SO(3)-reorientation of objects with a larger aspect ratio than in any prior work.
- We combine grasping and in-hand manipulation in a

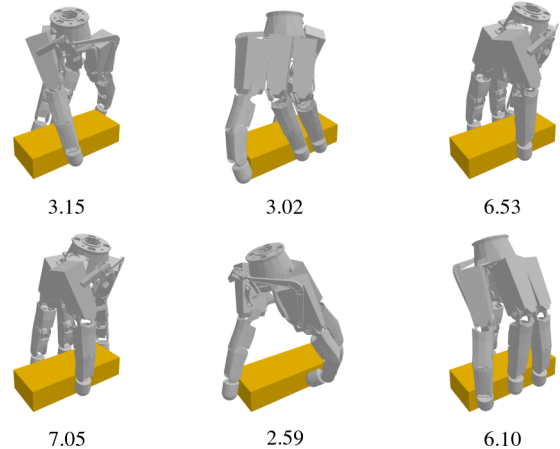


Fig. 2. A collection of different candidate grasps on the same object. The grasps shown were generated by our analytical grasp planner [1]. The grasp quality (measuring robustness) is shown below each grasp.

real-world pipeline on a humanoid robot, with generalization to objects unseen in training.

II. LEARNING DEXTROUS GRASPING AND IN-HAND MANIPULATION

In this section, we describe the two main building blocks of our system: the candidate grasp generation Section II-A and the in-hand manipulation agent Section II-B. The former generates a set of candidate grasps for a given object, while the latter is trained to reorient objects in-hand using only tactile feedback, i.e., position and torque readings. The methods described in this section are directly based on previous work on grasping [1, 40] and purely tactile in-hand manipulation [30, 2, 5, 30, 38] and described in more detail in the respective papers.

A. Candidate Grasp Generation

The initial grasping step can be defined as finding a set of K stable candidate grasps $\{[\mathcal{T}_H^i, q^i] \mid i = 1, \dots, K\}$ for a given object, with $\mathcal{T}_H^i$ being the hand pose and q^i being the joint configuration of the i -th grasp. Here, we make use of our learning-based grasping method described in Winkelbauer et al. [1] and Humt et al. [40]. This includes an analytical grasp planner which approaches grasping as an optimization problem with the objective being a grasp quality metric. Specifically, we base our metric on the epsilon quality metric with additional adjustments to consider the hand kinematics, independent force application per finger, and uncertainty in the relative pose between hand and object. Here, the quality metric can be interpreted as the minimum external wrench necessary to break the grasp, given that a fixed maximum torque budget is used on the finger joints. This means the planned grasps are not task-specific, but instead are optimized to handle all possible external disturbances equally well. See Fig. 2 for a random subset of grasps on the cuboid used throughout this work. Using

the analytical grasp planner, we generate a grasp dataset across different objects, which are then used to train our grasping network [1]. After training, the network can be used to efficiently generate arbitrary amounts of diverse grasps for a given object. The input to the network consists of the shape completed object observed only via a single depth image, which allows the grasp generation stage to generalize to unknown objects.

B. In-hand Manipulation Controller

In line with prior work on tactile in-hand manipulation [2, 5, 30, 38], we learn the in-hand reorientation controller by reinforcement learning in a realistic simulation environment. For training, we employ Estimator-Coupled Reinforcement Learning (ECRL) [30], concurrently learning a policy π and a state estimator f . The state estimator f recurrently estimates the object's pose $\hat{\mathcal{T}}_o^H$ from the history of joint measurements

$$\hat{\mathcal{T}}_o^H(t + \delta t) = f\left(\hat{\mathcal{T}}_o^H(t), z(t), \mathcal{S}\right), \quad (1)$$

where $\mathcal{T}_o^H$ denotes the object pose in hand frame H . The observation $z(t)$ is the concatenation of joint angle measurement q and control targets q_d stacked at a frequency of 60Hz for a window of 0.1s. The joint information enables the detection of contacts through an underlying high-fidelity impedance controller [41]. The shape encoding¹ $\mathcal{S}$ is obtained by transforming the canonical object mesh according to the estimated object pose $\hat{\mathcal{T}}_o^H$ and concatenating vectors between the object surface and a set of points fixed in H , as proposed in Pitz et al. [38]. We initialize the object pose at $t = 0$ with a known pose obtained from a vision system (see Section IV-B), subsequent estimates of object pose are done purely from tactile feedback. The estimator f is trained supervised to predict the object pose from sequences of joint measurements in simulation[30].

To achieve the in-hand reorientation, we train a goal-conditioned policy that outputs desired joint angles for the hand

$$q_d \sim \pi\left(\hat{R}_g^o, z, \mathcal{S}\right), \quad (2)$$

where $\hat{R}_g^o$ is the rotation to the goal relative to the currently estimated orientation. For learning the policy as well as the critic network, we use Proximal Policy Optimization (PPO)[42]. The reward function is described in Section III-C. Critically, during training, the hand object pose is initialized with a variety of grasps generated as described in Section II-A. Similar to prior work [30, 38], reorientations are considered successful if, after a time horizon of $\tau = 10s$, the object orientation is within a threshold $\theta = 0.4rad$ of the goal orientation. When the goal is reached, a new goal orientation is sampled uniformly from $SO(3)$, and the episode continues. Training is conducted in a simulation environment based on the Isaac Sim physics engine [43].

Suitable system identification and domain randomization of physical parameters (object masses, control gains, friction,

¹The shape encoding vector is computed in each timestep as a function of the estimates object pose $\hat{\mathcal{T}}_o^H$ and object mesh $\mathcal{M}$. For readability, we suppress this explicit dependence and write $\mathcal{S}$ instead of $\mathcal{S}(\hat{\mathcal{T}}_o^H, \mathcal{M})$.

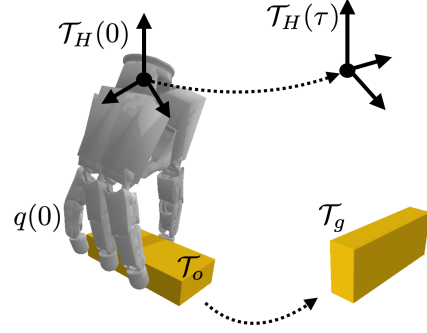


Fig. 3. Illustration of the problem setup. A multi-fingered hand is initially grasping an object (yellow) with pose $\mathcal{T}_o(0)$ using grasping joint angles $q(0)$ and is tasked with bringing it to a goal configuration $\mathcal{T}_g$ within time τ . The hand pose $\mathcal{T}_H$ is controlled by a robotic arm (not shown). When $\mathcal{T}_H$ is constrained due to the kinematics, the environment, or the task (e.g., the object needs to be placed on a table), arbitrary goal rotations of the object can only be achieved by in-hand manipulation. We are interested in finding $\mathcal{T}_H(0)$, $\mathcal{T}_H(\tau)$, and $q(0)$ that maximize the success probability of in-hand manipulation.

contact parametrization, and measurement noise) enables a robust sim2real transfer as validated in prior work [30, 38]. We refer to Röstel et al. [30] for an in-depth description of the ECRL training procedure and to Pitz et al. [38] for the shape-conditioned policy and estimator architectures.

After initialization with a known object pose, the resulting agent is able to reorient objects without external supervision of the object pose, purely from tactile feedback (i.e., torque and position readings). When autonomously deploying the agent, the policy is stopped when the estimated object orientation $\hat{R}_o^H$ is within a threshold of the goal orientation R_g^H .

III. COMPOSING DEXTROUS GRASPING AND IN-HAND MANIPULATION

This section outlines the core contribution of this paper: a novel method for maximizing the end-to-end performance of a grasp-manipulation pipeline. We first formulate an optimization problem over initial hand-object configurations, i.e., dextrous grasps. We then show how the separately trained grasp generator and in-hand manipulation agent from Section II can be composed by scoring the generated grasps using the critic network trained for in-hand manipulation.

A. Problem Formulation

We study the problem of bringing an object with initial pose $\mathcal{T}_o(t = 0) = [x_o(0), R_o(0)] \in SE(3)$ to a goal configuration $\mathcal{T}_g = [x_g, R_g] \in SE(3)$ in world frame using a multi-fingered hand mounted on a robotic arm (Fig. 3). While the position component x can be directly controlled by the robot arm to match x_g , the orientation component is typically not directly controllable and is most efficiently achieved by in-hand manipulation. For this, we use an in-hand reorientation controller (see Section II-B) that reorients an object from the initial object orientation $R_o^H(0)$ to R_g^H in

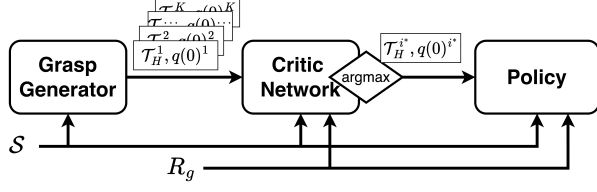


Fig. 4. Scheme of the proposed method for selecting suitable grasps for in-hand manipulation. First, a set of candidate grasps is generated for a given object shape $\mathcal{S}$ and pose $\mathcal{T}_o$. The grasp candidates $[\mathcal{T}_H^i, q^i]$ are generated w.r.t stability and satisfy applicable task constraints such as reachability and collision avoidance. Given a desired goal orientation R_g , the critic network is then used to score the grasps based on their suitability for achieving the desired rotation in-hand. The highest-scoring grasp is then executed and used to initialize the in-hand manipulation policy.

hand frame H within time $t = \tau$ with success probability

$$P(\text{success} | R_o^H(0), R_g^H, q(0), \mathcal{S}), \quad (3)$$

where “success” means that the angle between object orientation and goal orientation $d(R_o^H(\tau), R_g^H)$ is below a threshold θ . In (3), the probability of success additionally depends on the initial joint angles of the hand $q(0)$ and the object’s shape $\mathcal{S}$. The probabilistic formulation is motivated by the presence of inherent uncertainties in the system, e.g., due to sensor noise, unknown physical properties of the object and incomplete state information.

We are now interested in finding initial and final hand poses $\mathcal{T}_H(0)$ and $\mathcal{T}_H(\tau)$, as well as grasping joint angles $q(0)$ that maximize (3) subject to $\mathcal{T}_o(\tau) = \mathcal{T}_g$. Without loss of generality, for the experiments in this paper, we only consider hand poses $\mathcal{T}_H$ such that the hand is pointing downwards, which is amenable to grasping from and placing on flat surfaces. Additionally, we only allow rotating the hand base by at most $\pi/2$ around the vertical axis to avoid twisted arm configurations and optionally impose a fixed wrist-rotation constraint $\mathcal{T}_H(0) = \mathcal{T}_H(\tau)$.

In this work, we require visual information only for inferring the initial state $\mathcal{T}_o(0)$ (and possibly the object’s shape $\mathcal{S}$); for subsequent grasping and manipulation, we rely purely on tactile (i.e., torque-sensing) feedback of the hand.

B. Optimizing Over Initial States By Critic-Scoring

In reinforcement learning, the value function $V(s_t)$ is defined as the expected discounted sum of rewards r_t starting from state s_t and following the policy π thereafter

$$V(s_t) = \mathbb{E}_\pi \left[\sum_{k=t}^{\infty} \gamma^{k-t} r_k \right] \quad (4)$$

with discount factor γ . In actor-critic methods like PPO, the value function is approximated by a neural network $v(s_t) \approx V(s_t)$ and learned concurrently with the policy by minimizing a temporal difference loss (see, e.g., [44]). We note that for a sparse reward of the form

$$r_s(t) = \begin{cases} 1 & \text{if } t = \tau \text{ and success} \\ 0 & \text{otherwise,} \end{cases} \quad (5)$$

a well-motivated assumption is that the value at $t = 0$

$$V(t = 0) = \gamma^T \mathbb{E}_\pi [r_s(t = \tau)], \quad (6)$$

is a good proxy for the success probability of a given initial task configuration with normalization factor γ^T .

In the context of optimization over initial states $s(0)$, we, therefore, propose to use the learned critic network (as obtained from learning the in-hand manipulation policy in Section II-B) for choosing the most suitable initial state for the task at hand. A benefit of this post hoc optimization approach is that the initial states s_i to be evaluated (referred to as *candidate states*) can be chosen to match specific task constraints, which we will exploit in our application to grasp selection. Moreover, evaluating batches of initial states is computationally efficient as it requires only a forward pass through the critic network.

C. Scoring Grasps for In-Hand Manipulation

We now apply the critic-scoring approach described above to the problem of selecting suitable grasps for in-hand manipulation. For learning in-hand manipulation, we use a reward composed of a dense and a sparse component $r = r_d + r_s$. The sparse reward is given after successful in-hand reorientation. Because training with only sparse rewards is challenging, we also use a dense reward that encourages rotating the object towards the goal while penalizing deviations in object position and joint position from their default values (see Pitz et al. [38]). In practice, we parametrize the critic v as a single Multi-Layer-Perceptron (MLP) with a two-dimensional output (v_d, v_s) estimating (4) for the dense and sparse reward terms separately. While both outputs can be used for training the policy, only v_s is used for scoring the initial states as a proxy for expected success (6). The input to the critic network is the same as for the policy (2).

Given a set of candidate grasps $\{[\mathcal{T}_H^i, q^i] \mid i = 1, \dots, K\}$ for initial object pose $\mathcal{T}_o(0)$ and a goal orientation R_g , we would like to find the most suitable grasps for subsequent in-hand manipulation, i.e., that maximizes (3). To this end, we score each grasp by evaluating the sparse component of the critic network v_s for the initial state and choose the highest-scoring candidate

$$i^* = \arg \max_i v_s \left(\hat{R}_g^{o,i}(0), z^i(0), \mathcal{S} \right), \quad (7)$$

where the object pose $\mathcal{T}_o^{H,i}$ and goal rotation $R_g^{H,i}$ are obtained by transformation of $\mathcal{T}_o$ and R_g to the hand frame $\mathcal{T}_H^i$. Note that because the policy and critic expect inputs in hand frame H , the critic is agnostic to absolute hand-pose transformations. This opens the possibility for separate optimization over initial and final hand poses $\mathcal{T}_H(0)$ and $\mathcal{T}_H(\tau)$ respectively (*move base*), effectively changing only the goal orientation in hand frame R_g^H for the same goal orientation in world frame R_g .

IV. EXPERIMENTS

In this section, we evaluate the proposed method in simulation and on a real robot. We assess the effect of grasp scoring

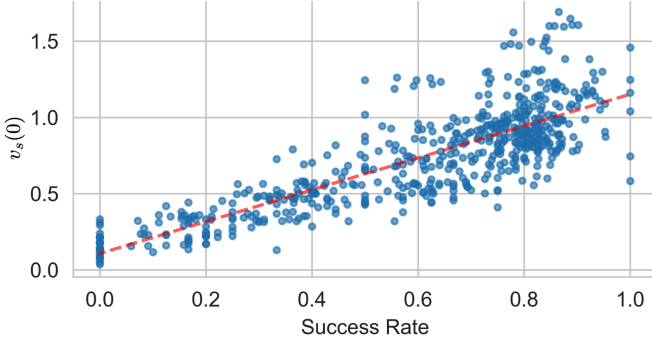


Fig. 5. Correlation between critic score $v_s(t = 0)$ and success rate of in-hand manipulation at timestep τ for the long cuboid object in simulation. Each data point corresponds to a pair of initial object rotation and goal rotation $(R_o(0), R_g)$, where $R_o(0)$ are aggregated within 0.3rad to calculate the success rate.

on two in-hand manipulation policies: first, a shape-conditioned policy that is trained to reorient randomized geometric shapes, generalizing to new shapes at test time similar to Pitz et al. [38]. Second, we train a policy to reorient a cuboid with dimensions 4cm $\times$ 8cm $\times$ 20cm, which represents to our knowledge the largest aspect ratio that has been considered for general $SO(3)$ in-hand reorientation.

For a fair comparison to baselines, we only consider *stable* initial grasps when assessing their suitability for in-hand reorientation.

A. Evaluation in Simulation

In simulation, we conduct a series of evaluations to assess the effectiveness of the critic-scoring approach proposed in Section III. For the experiments described in this section, we use a separate set of candidate grasps different from those used for training and leave domain randomization enabled during evaluation.

1) **Correlation of Critic Score and Success Rate:** We first show that the output of the critic network is predictive of in-hand manipulation success. In Fig. 5, we compare the value of $v_s(t = 0)$ to the success rate of in-hand manipulation at timestep τ for various initial and goal rotations of the long cuboid object.

The data shows a clear correlation between the critic score and success rate with a Pearson correlation coefficient of 0.82. We attribute the variance in the data mainly to the effect of noise injected into the simulation (domain randomization).

Note that the values of v_s are not directly interpretable as probabilities here, even after normalization. This is expected, as the critic also takes into account the value of reaching goals after the horizon τ at multiples of τ thereafter (see Section II-B). Still, we expect, up to approximation errors and unobserved states, a monotonic relation between the success rate and the critic score (6), justifying the use of the critic network for scoring grasps.

2) **Effect of Selecting Grasps by Critic Score:** We now assess the effect of grasp optimization on the in-hand manipulation performance. To this end, we consider the following

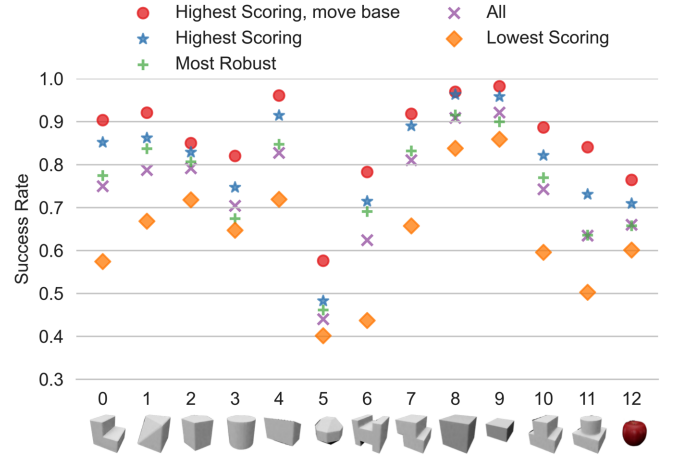


Fig. 6. Success rates of in-hand manipulation in simulation for different strategies of choosing the initial grasp. Objects from Pitz et al. [38], with objects 0-7 being in the training distribution and objects 8-12 out-of-distribution for the shape-conditioned policy. For each object, we simulate $N \approx 50k$ trials, leading to negligible statistical error.

strategies for selecting the initial grasps among candidate grasps for a given object pose and goal orientation:

- **Most Robust:** Select the grasp with the highest epsilon quality metric.
- **Highest Scoring:** Select the grasp with the highest critic score (7), i.e., optimization over $q(0)$ and $\mathcal{T}_H(0) = \mathcal{T}_H(\tau)$.
- **Highest Scoring, move base:** Select the grasp with the highest critic score while allowing the hand base to rotate around the vertical axis by at most $\pi/2$, i.e., in general $\mathcal{T}_H(0) \neq \mathcal{T}_H(\tau)$.
- **Lowest Scoring:** Select the grasp with the lowest critic score (without base movement).
- **All:** Average across all (stable) grasp candidates.

Note that the object rotation in world frame that needs to be accomplished by in-hand manipulation is unaffected by choice of grasp, except in the case of *Highest Scoring, move base*.

In figure Fig. 6, we show in-hand manipulation success rates for different strategies of selecting the initial grasp. Across different objects, we find that selecting grasps based on the critic score consistently increases the success rate of in-hand manipulation compared to the average grasp or selecting grasps based on the epsilon quality metric. Allowing the hand to rotate around the vertical axis as preferred by the critic network further increases the success rate. In Table I, we additionally report the improvements in success rate, the fraction of trials where the object is dropped during manipulation, and the time to reach the goal orientation. We aggregate all quantities over objects for the shape-conditioned policy with $N \approx 600k$ trials, leading to high statistical significance. While selecting the highest-scoring grasp already leads to a moderate decrease in both the fraction of objects dropped and runtime, the most prominent improvement is again observed when additionally allowing

TABLE I
EFFECT OF GRASP SELECTION STRATEGY ON IN-HAND MANIPULATION PERFORMANCE IN SIMULATION

Grasp Selection Strategy	Success Rate Improvement [%] $\uparrow$	Object Dropped [%] $\downarrow$	Average Time to Goal[s] $\downarrow$
All	-	3.8	4.2
Most Robust	+2.3	2.9	4.2
Highest-Scoring	+9.4	2.5	3.6
Highest-Scoring, move base	+17.6	1.9	2.7
Lowest-Scoring	-14.6	7.8	4.7

the hand base to rotate (*Highest-Scoring, move base*). This can in-parts be explained by the ability of this strategy to partially compensate for in-hand reorientation by rotating the hand base (by at most $\pi/2$), which we find leads to an average decrease in required in-hand reorientation angle of approx. 0.4 rad. For the large cuboid object, we observe a similar trend, with success rates of 78% for the *Most Robust*, 84% for the *Highest Scoring*, and 90% for the *Highest Scoring, move base* strategies.

B. Real-World Validation

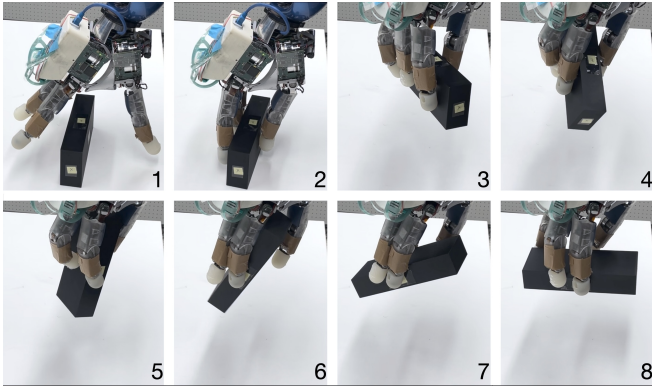


Fig. 7. Real-world grasp-manipulation sequence for the large cuboid object. 1) Pre-grasping 2) Grasping pose with highest-scoring joint angles $q(0)$ and hand pose $\mathcal{T}_H(0)$ 3) Arm lifts the object from the table and rotates to final hand pose $\mathcal{T}_H(\tau)$ (as predicted by critic scoring) without finger joint movement 4)-7) In-hand manipulation by policy without supporting surface 8) Goal orientation reached.

The combination of grasping and in-hand manipulation is evaluated on our precisely calibrated [45, 46] humanoid robot *Agile Justin* [47]. The given object is observed as an incomplete point cloud using a Kinect depth camera [48]. Using our object-agnostic shape completion [1], a full 3D model is reconstructed and used to estimate the initial pose $\mathcal{T}_o(0)$ of the object. This pose estimation is done via a global estimator using RANSAC registration based on FPFH feature matching [49], followed by a refinement stage based on the iterative closest point algorithm [50].

Next, a set of 200 grasps $[\mathcal{T}_H^i(0), q^i]$ is predicted by our grasping network using the shape-completed 3D model and filtered based on collisions with the environment. Additionally, 90 reachable target hand poses $\mathcal{T}_H(\tau)$ are generated. For each of the up to 18 000 combinations, the critic network v_s is evaluated, and the candidate combination i^* of initial hand

pose, final hand pose, and grasping joint configuration with maximum value is selected. Evaluating the critic network for a batch of 18 000 inputs takes ~ 0.1 s on a T4 GPU, including the processing of the shape encoding.

Using a learning-based motion planner [51], the hand now approaches the object as specified by the selected grasp $[\mathcal{T}_H^{i^*}, q^{i^*}]$. Afterward, the fingers are closed until they contact the object, and the hand is lifted up into the final hand pose $\mathcal{T}_H^{i^*}(\tau)$.

Now, the in-hand manipulation using the policy π is started to reorient the object towards the given goal rotation R_g^{H,i^*} , without tracking of the object pose from the vision system (which would be impractical due to occlusions). The policy stops autonomously when the object orientation $\hat{R}_o^H$ estimated from tactile feedback is within a threshold of the goal orientation R_g^H . If the desired goal orientation was actually approximately reached, the trial is marked as a success. In this procedure, two different objects are evaluated: the large cuboid, shown in Fig. 7, and an object that is out-of-distribution for the shape-conditioned policy, both shown in Fig. 1. For both objects, we exhaustively evaluate 24 goal rotations in a $\pi/2$ -discretization of $SO(3)$ (octahedral group), with 22 successful reorientations for the cuboid and 23 successes for the OOD object, representing an aggregated success rate of 93.7%. Real-world rollouts for these and more objects can be found in the supplementary video.

V. CONCLUSION

In this work, we combine dexterous grasping and goal-conditioned in-hand object reorientation with a multi-fingered hand. As the core idea, we propose the use of the critic network to choose grasps that maximize the probability of successful in-hand manipulation. Our experiments showed that the proposed method substantially increases the success rate of in-hand manipulation compared to selecting grasps based on robustness alone. We further demonstrated that the predicted values can be used to decide which movement of the robotic arm promotes the reorientation task.

A limitation of the current implementation is that the initial pose estimate obtained from vision can be ambiguous (e.g. due to occlusions). In the future, we want to resolve such uncertainty by using tactile exploration during in-hand manipulation. Another interesting direction for future work is to balance different objectives when optimizing grasps, including robustness and critic scores, to allow a more fine-grained, task-specific trade-off.

REFERENCES

- [1] D. Winkelbauer *et al.*, “A two-stage learning architecture that generates High-Quality grasps for a Multi-Fingered hand,” in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2022.
- [2] L. Sievers, J. Pitz, and B. Bäuml, “Learning purely tactile in-hand manipulation with a torque-controlled hand,” in *Proc. IEEE International Conference on Robotics and Automation*, 2022.
- [3] OpenAI *et al.*, “Learning dexterous In-Hand manipulation,” *arXiv preprint*, 2018.
- [4] T. Chen, J. Xu, and P. Agrawal, “A system for general in-hand object re-orientation,” *Conference on Robot Learning*, 2021.
- [5] J. Pitz, L. Röstel, L. Sievers, and B. Bäuml, “Dextrous tactile In-Hand manipulation using a modular reinforcement learning architecture,” in *Proc. IEEE International Conference on Robotics and Automation*, 2023.
- [6] J. Butterfaß, M. Grebenstein, H. Liu, and G. Hirzinger, “DLR-Hand II: Next generation of a dextrous robot hand,” in *Proc. IEEE International Conference on Robotics and Automation*, 2001.
- [7] V.-D. Nguyen, “Constructing Force-Closure Grasps,” *International Journal of Robotics Research*, 1988.
- [8] Z. Li and S. Sastry, “Task-oriented optimal grasping by multifingered robot hands,” *IEEE Journal on Robotics and Automation*, vol. 4, no. 1, 1988.
- [9] D. G. Kirkpatrick, B. Mishra, and C. K. Yap, “Quantitative Steinitz’s theorems with applications to multi-fingered grasping,” in *ACM Symposium on Theory of Computing*, 1990.
- [10] C. Ferrari and J. Canny, “Planning optimal grasps,” in *IEEE International Conference on Robotics and Automation*, 1992.
- [11] R. Pelossof, A. Miller, P. Allen, and T. Jebara, “An SVM learning approach to robotic grasping,” in *IEEE International Conference on Robotics and Automation*, 2004.
- [12] A. Saxena, J. Driemeyer, J. Kearns, and A. Y. Ng, “Robotic Grasping of Novel Objects,” *Advances in neural information processing systems*, vol. 19, 2006.
- [13] I. Lenz, H. Lee, and A. Saxena, “Deep learning for detecting robotic grasps,” *International Journal of Robotics Research*, vol. 34, no. 4-5, 2015.
- [14] J. Varley, J. Weisz, J. Weiss, and P. Allen, “Generating multi-fingered robotic grasps via deep learning,” in *IEEE International Conference on Intelligent Robots and Systems*, 2015.
- [15] D. Kappler, J. Bohg, and S. Schaal, “Leveraging big data for grasp planning,” in *IEEE International Conference on Robotics and Automation*, 2015.
- [16] Y. Li *et al.*, “HGC-Net: Deep Anthropomorphic Hand Grasping in Clutter,” in *IEEE International Conference on Robotics and Automation*, 2022.
- [17] W. Wei *et al.*, “DVGG: Deep Variational Grasp Generation for Dextrous Manipulation,” *IEEE Robotics and Automation Letters*, vol. 7, no. 2, 2022.
- [18] M. A. Roa and R. Suárez, “Grasp quality measures: review and performance,” *Autonomous Robots*, vol. 38, no. 1, 2015.
- [19] J. K. Salisbury and J. J. Craig, “Articulated Hands: Force Control and Kinematic Issues,” *International Journal of Robotics Research*, vol. 1, no. 1, 1982.
- [20] C. A. Klein and B. E. Blaho, “Dexterity Measures for the Design and Control of Kinematically Redundant Manipulators,” *International Journal of Robotics Research*, vol. 6, no. 2, 1987.
- [21] T. Yoshikawa, “Manipulability of Robotic Mechanisms,” *International Journal of Robotics Research*, vol. 4, no. 2, 1985.
- [22] J. T. Y. Wen and L. S. Wilfinger, “Kinematic manipulability of general constrained rigid multibody systems,” *IEEE Transactions on Robotics and Automation*, vol. 15, no. 3, 1999.
- [23] D. Prattichizzo, M. Malvezzi, M. Gabiccini, and A. Bicchi, “On the manipulability ellipsoids of underactuated robotic hands with compliance,” *Robotics and Autonomous Systems*, vol. 60, no. 3, 2012.
- [24] M. Sato and T. Yoshikawa, “A grasp performance criterion for robot hands considering multiple aspects of tasks and hand configurations,” in *2011 IEEE International Conference on Robotics and Biomimetics*, 2011.
- [25] A. Okamura, N. Smaby, and M. Cutkosky, “An overview of dextrous manipulation,” in *Proc. IEEE International Conference on Robotics and Automation*, 2000.
- [26] J.-P. Saut, A. Sahbani, S. El-Khoury, and V. Perdereau, “Dextrous manipulation planning using probabilistic roadmaps in continuous grasp subspaces,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2007.
- [27] A. S. Morgan *et al.*, “Complex in-hand manipulation via compliance-enabled finger gaiting and multi-modal planning,” *IEEE Robotics and Automation Letters*, 2022.
- [28] H. Qi *et al.*, “In-Hand object rotation via rapid motor adaptation,” *Proc. Conference on Robot Learning*, 2022.
- [29] G. Khandate *et al.*, “Sampling-based exploration for reinforcement learning of dextrous manipulation,” *arXiv preprint*, 2023.
- [30] L. Röstel, J. Pitz, L. Sievers, and B. Bäuml, “Estimator-Coupled reinforcement learning for robust purely tactile In-Hand manipulation,” in *Proc. IEEE-RAS 22nd International Conference on Humanoid Robots*, 2023.
- [31] OpenAI *et al.*, “Solving rubik’s cube with a robot hand,” *arXiv preprint*, 2019.
- [32] A. Handa *et al.*, “Dextreme: Transfer of agile in-hand manipulation from simulation to reality,” *Proc. IEEE International Conference on Robotics and Automation*, 2022.
- [33] G. Khandate *et al.*, “R $\times$ r: Rapid exploration for re-

- inforcement learning via sampling-based reset distributions and imitation pre-training,” *Autonomous Robots*, vol. 48, no. 7, 2024.
- [34] T. Chen *et al.*, “Visual dexterity: In-hand reorientation of novel and complex object shapes,” *Science Robotics*, 2022.
- [35] H. Qi *et al.*, “General In-Hand object rotation with vision and touch,” *Proc. Conference on Robot Learning*, 2023.
- [36] Y. Yuan *et al.*, “Robot synesthesia: In-hand manipulation with visuotactile sensing,” *arXiv preprint*, 2023.
- [37] J. Wang *et al.*, “Lessons from learning to spin ”pens”,” *arXiv preprint*, 2024.
- [38] J. Pitz *et al.*, “Learning a shape-conditioned agent for purely tactile in-hand manipulation of various objects,” *arXiv*, 2024.
- [39] M. B. Horowitz and J. W. Burdick, “Combined grasp and manipulation planning as a trajectory optimization problem,” in *IEEE International Conference on Robotics and Automation*, 2012.
- [40] M. Humt, D. Winkelbauer, U. Hillenbrand, and B. Bäuml, “Combining shape completion and grasp prediction for fast and versatile grasping with a multi-fingered hand,” in *Proc. IEEE-RAS 22nd International Conference on Humanoid Robots*, 2023.
- [41] L. Röstel, L. Sievers, J. Pitz, and B. Bäuml, “Learning a state estimator for tactile In-Hand manipulation,” in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2022.
- [42] J. Schulman *et al.*, “Proximal policy optimization algorithms,” *arXiv preprint arXiv:1707.06347*, 2017.
- [43] NVIDIA, “Omniverse isaac gym reinforcement learning environments for isaac sim,” <https://github.com/NVIDIA-Omniverse/OmnIsaacGymEnvs>, July 2023.
- [44] J. Schulman *et al.*, “High-dimensional continuous control using generalized advantage estimation,” *arXiv [cs.LG]*, 2015.
- [45] J. Tenhumberg, D. Winkelbauer, D. Burschka, and B. Bäuml, “Self-contained calibration of an elastic humanoid upper body using only a head-mounted rgb camera,” in *IEEE International Conference on Humanoid Robots*, 2022.
- [46] J. Tenhumberg, L. Sievers, and B. Bäuml, “Self-contained and automatic calibration of a multi-fingered hand using only pairwise contact measurements,” in *Proc. IEEE-RAS 22nd International Conference on Humanoid Robots*, 2023.
- [47] B. Bäuml *et al.*, “Agile justin: An upgraded member of DLR’s family of lightweight and torque controlled humanoids,” in *Proc. IEEE International Conference on Robotics and Automation*, 2014.
- [48] R. Wagner, U. Frese, and B. Bäuml, “Real-time dense multi-scale workspace modeling on a humanoid robot,” in *IEEE International Conference on Intelligent Robots and Systems*, 2013.
- [49] R. B. Rusu, N. Blodow, and M. Beetz, “Fast point feature histograms (fpfh) for 3d registration,” in *2009 IEEE International Conference on Robotics and Automation*, 2009.
- [50] P. Besl and N. D. McKay, “A method for registration of 3-d shapes,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, 1992.
- [51] J. Tenhumberg, D. Burschka, and B. Bäuml, “Speeding up optimization-based motion planning through deep learning,” in *IEEE International Conference on Intelligent Robots and Systems*, 2022.