On the Feasibility of Learning Finger-gaiting In-hand Manipulation with Intrinsic Sensing

Gagan Khandate, Maximilian Haas-Heger, Matei Ciocarlie

***Abstract*— Finger-gaiting manipulation is an important skill to achieve large-angle in-hand re-orientation of objects. However, achieving these gaits with arbitrary orientations of the hand is challenging due to the unstable nature of the task. In this work, we use model-free reinforcement learning (RL) to learn finger-gaiting only via precision grasps and demonstrate finger-gaiting for rotation about an axis purely using on-board proprioceptive and tactile feedback. To tackle the inherent instability of precision grasping, we propose the use of initial state distributions that enable effective exploration of the state space. Our method can learn finger gaiting with significantly improved sample complexity than the state-of-the-art. The policies we obtain are both robust and generalizable to novel objects.**

I. Introduction

Dexterous in-hand manipulation [1] is the ability to move a grasped object from one pose to another desired pose. Humans routinely use in-hand manipulation to perform tasks such as re-orienting a tool from its initial grasped pose to a useful pose, securing a better grasp on the object, exploring the shape of an unknown object, etc. Thus, robotic in-hand manipulation is an important step towards the general goal of manipulating objects in cluttered and unstructured environments such as in a kitchen or a warehouse. However, versatile in-hand manipulation remains a long standing challenge.

A whole spectrum of methods have been considered for in-hand manipulation; online trajectory optimization methods [2] and model-free deep reinforcement learning (RL) methods [3] stand out for highly actuated dexterous hands. Model-based online trajectory optimization methods have succeeded in generating complex behaviors for dexterous robotic manipulation, but not for finger-gaiting as these tasks fatally exacerbate their limitations: transient contacts introduce large non-linearities in the model, which also depends on hard-to-model contact properties.

While RL has been successful in demonstrating diverse in-hand manipulation skills both in simulation and on real hands [4], the policies obtained are object centric and require large training times. In some cases, these policies have not been demonstrated with arbitrary orientations of the hand as they expect the palm to support the object during manipulation —a consequence of the policies being trained with the hand in palm-up orientation which simplifies training. In other cases, policies require extensive external sensing involving multi-camera systems to track the fingers and/or the object, systems that are hard to deploy outside the lab environments.

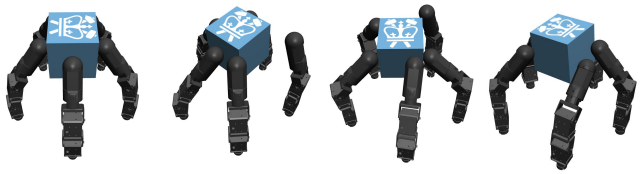


Fig. 1: A learned finger-gaiting policy that can continuously re-orient the target object about the hand z-axis. The policy only uses sensing modalities intrinsic to the hand (such as touch and proprioception), and does not require explicit object pose information from external sensors.

Tactile feedback has strong potential in enabling robust and generalizable in-hand manipulation skills [5] and in eliminating the need for external sensing. However, integrating tactile feedback with RL is a challenge on its own. Besides the general difficulty of simulating the transduction involved, tactile feedback is often high dimensional which can prohibitively drive up the number of training samples required. Hence, prior work using RL for in-hand manipulation either totally avoid using tactile feedback or consider tasks requiring fewer training samples where it is feasible to learn directly with the real hand.

We too use RL, but focus on learning finger-gaiting (manipulation involving finger substitution and re-grasping) and finger-pivoting (manipulation involving the object in hinge-grasp) skills. Both skills are important towards enabling arbitrary large-angle in-hand object re-orientation: achieving an indefinitely-large rotation of the grasped object around a given axis, up to or even exceeding a full revolution. Such a task is generally not achievable by in-grasp manipulation (i.e. without breaking the contacts of the original grasp) and requires finger-gaiting or finger-pivoting (i.e. breaking and re-establishing contacts during manipulation); these are not restricted by the kinematic constraints of the hand and can achieve potentially limitless object re-orientation.

We are interested in achieving these skills exclusively through using fingertip grasps (i.e precision in-hand manipulation [6]) without requiring the presence of the palm underneath the object, which enables the policies to be used in arbitrary orientations of the hand. However, the task of learning to manipulate only via such precision grasps is a significantly harder problem: action randomization, responsible for exploration in reinforcement learning, often fails as the hand can easily drop the object.

Gagan Khandate is with Department of Computer Science, Columbia University, New York, USA, gagank@cs.columbia.edu

Maxmillian Haas-Heger is with Nuro, Mountain View, CA, USA, m.haas@columbia.edu

Matei Ciocarlie is with Department of Mechanical Engineering, Columbia University, New York, USA, matei.ciocarlie@columbia.edu

Furthermore, we would like to circumvent the need for cumbersome external sensing by only using internal sensing in achieving these skills. The challenge here is that the absence of external sensing implies we do not have information regarding the object such as its shape and pose. On the other hand, internal sensing by itself can provide object information sufficient towards our goal.

We set out to determine if we can even achieve finger-gaiting and finger-pivoting skills purely through intrinsic sensing in simulation, where we evaluate both proprioceptive feedback and tactile feedback. To this end, we consider continuous object re-orientation about a given axis towards learning finger-gaiting and finger-pivoting without object pose information. With this approach, we hope to learn policies to rotate object about cardinal axes and combine them for arbitrary in-hand object re-orientation. To overcome challenges in exploration, we propose collecting training trajectories starting from a wide range of grasps sampled from appropriately designed initial state distributions as an alternative exploration mechanism.

We summarize the contributions of this work as follows:

1) We learn finger-gaiting and finger-pivoting policies that can achieve large angle in-hand re-orientation of a range of simulated objects. Our policies learn to grasp and manipulate only via precision fingertip grasps using a highly dexterous and fully actuated hand, allowing us to keep the object in a stable grasp without the need for passive support at any instance during manipulation.
2) We are the first to achieve these skills while only making use of intrinsic sensing such as proprioception and touch, while also generalizing to multiple object shapes.
3) We present an exhaustive analysis of the importance of different internal sensor feedback for learning finger-gaiting and finger-pivoting policies in a simulated environment using our approach.

II. Related Work

Early model-based work on finger-gaiting [7][8] [9] [10] and finger-pivoting [11] generally make simplifying assumptions such as 2D manipulation, accurate models, and smooth object geometries which limit their versatility. More recently, Fan et al. [12] and Sundaralingam et al. [13] use model based online optimization and demonstrate finger-gaiting in simulation. These methods either use smooth objects or require accurate kinematic models of the of the object, which make these methods challenging to transfer to real hands.

OpenAI et al. [4] demonstrate finger-gaiting and finger-pivoting using RL, but as previously discussed, their policies cannot be used for arbitrary orientations of the hand. This can be achieved using only force-closed precision fingertip grasps, but learning in-hand manipulation using only these grasps is challenging with few prior work. Li et al. [14] learn 2D re-orientation using model-based controllers to ensure grasp stability in simulation. Veiga et al. [15] demonstrate in-hand reorientation with only fingertips but these object centric policies are limited to small re-orientations via in-grasp manipulation and still require external sensing. Shi et al. [16] demonstrate precision finger-gaiting but only on a lightweight ball. Morgan et al. [17] also show precision finger-gaiting but with an under-actuated hand specifically designed for this task. We consider finger-gaiting with a highly actuated hand; our problem is exponentially harder due to increased degrees of freedom leading to poor sample complexity.

Some prior work [18][19][20] use human expert trajectories to improve sample complexity for dexterous manipulation. However, these expert demonstrations are hard to obtain for precision in-hand manipulation tasks and even more so for non-anthropomorphic hands. Alternatively, model-based RL has also been considered for some in-hand manipulation tasks: Nagabandi et al. [21] manipulate boading balls but use the palm for support; Morgan et al. [17] learn finger-gaiting but with a task specific underactuated hand. However, learning a reliable forward model for precision in-hand manipulation with a fully dexterous hand can be challenging. Collecting data involves random exploration, which, as we discuss later in this paper, has difficulty exploring in this domain.

Prior work using model-free RL for manipulation rarely use tactile feedback as tactile sensing available on real hand is often high dimensional and hard to simulate [4]. Hence, van Hoof et al. [22] propose learning directly on a real hand, but this naturally limits us to tasks learnable on real hands. Alternatively, Veiga et al. [15] learn a higher level policy through RL, while having low level controllers exclusively deal with tactile feedback. However, this method deprives the policy from leveraging tactile feedback beneficial in other challenging tasks. While Melnik et al. [23] show that using tactile feedback improves sample complexity in such tasks, they use high-dimensional tactile feedback with full coverage of the hand that is hard to replicate on a real hand. We instead consider low-dimensional tactile feedback covering only the fingertips.

Contemporary to our work, Chen et al. [24] show in-hand re-orientation without support surfaces that generalizes to novel objects. The policies exhibit complex dynamic behaviors including occasionally throwing the object and re-grasping it in the desired orientation. We differ from this work as our policies only use sensing that is internal to the hand, and always keep the object in a stable grasp to be robust to perturbation forces at all times. Furthermore, our policies require a number of training samples that is smaller by multiple orders of magnitude, a feature that we attribute to efficient exploration via appropriate initial state distributions.

III. Learning precision in-hand re-orientation

In this work, we address two important challenges for precision in-hand re-orientation using reinforcement learning. First, we propose a hand-centric decomposition method for achieving arbitrary in-hand re-orientation in an object-agnostic fashion. Next, we identify that a key challenge

Fig. 2: Hand-centric decomposition of in-hand re-orientation into re-orientation about cardinal axes

of exploration for learning precision in-hand manipulation skills can be alleviated by collecting training trajectories starting at varied stable grasps. We use these grasps to design appropriate initial state distributions for training. Our approach assumes a fully-actuated and position-controlled hand.

A. Hand-centric decomposition

Our aim is to push the limits on manipulation with only intrinsic sensing, and do this in a general fashion without assuming object knowledge. Thus, we do so in a hand-centric way: we learn to rotate around axes grounded in the hand frame. This means we do not need external tracking (which presumably needs to be trained for each individual object) to provide object-pose[1]. We also find that rewarding angular velocity about desired axis of rotation is conducive to learning finger-gaiting and finger-pivoting policies. However, learning a single policy for any arbitrary axis is challenging as it involves learning goal-conditioned policies, which is difficult for model free RL.

Our proposed method for wide arbitrary in-hand re-orientation is thus to decompose the problem of achieving arbitrary angular velocity of the object into learning separate policies about the cardinal axes as shown in Fig. 2. The finger-gaiting policies obtained for each axis can then be combined in the appropriate sequence to achieve the desired change in object orientation, while side-stepping the difficulty of learning a goal-conditioned policy.

We assume proprioceptive sensing can provide current positions $\boldsymbol{q}$ and controller set-point positions $\boldsymbol{q}_d$. We note that the combination of desired positions and current positions can be considered as a proxy for motor forces, if the characteristics of the underlying controller are fixed. More importantly, we assume tactile sensing to provide contact positions $\boldsymbol{c}^i$ and normal forces t_n^i on each fingertip i. With known fingertip geometry, the contact normals $\hat{\mathbf{t}}_n^i$ can be derived from contact positions $\boldsymbol{c}^i$.

Our axis-specific re-orientation policies are conditioned only on proprioceptive and tactile feedback as given by the observation vector $\mathbf{o}$:

$$\mathbf{o} = [\boldsymbol{q}, \boldsymbol{q}_d, \boldsymbol{c}^1 \dots \boldsymbol{c}^m, t_n^1 \dots t_n^m, \hat{\mathbf{t}}_n^1 \dots \hat{\mathbf{t}}_n^m] \tag{1}$$

Our policies command set-point changes $\Delta\mathbf{q}_d$ which we henceforth refer to by action $\mathbf{a}$.

[1]We note that there exist applications where specific object poses are needed, and for such cases we envision future work where a high-level object-specific tracker makes use of our hand-centric object-agnostic policies to achieve it.

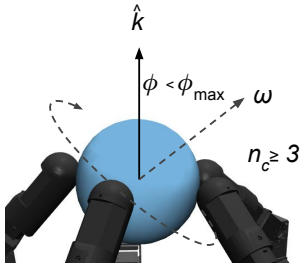


Fig. 3: Learning axis conditional continuous re-orientation $\hat{\boldsymbol{k}}$. We use the component of angular velocity $\boldsymbol{\omega}$ about $\hat{\boldsymbol{k}}$ as reward when we the object is a grasp with 3 or more fingertips, i.e $n_c \geq 3$.

B. Learning axis-specific re-orientation

We now describe the procedure for learning in-hand re-orientation policies for an arbitrary but fixed axis. Let $\hat{\boldsymbol{k}}$ be the desired axis of rotation. To learn policy axis-specific policy $\pi^{\hat{\boldsymbol{k}}}$ that continuously re-orients the object about the desired axis we use the object's angular velocity $\boldsymbol{\omega}$ along $\hat{\boldsymbol{k}}$ as reward as shown in Fig 3. However, to ensure that the policy learns to only use precision fingertip grasps to re-orient the object, we provide this reward if only fingertips are in contact with the object. In addition, we require that at least 3 fingertips are in contact with the object as they can achieve force closure. Also, we encourage alignment of the object's axis of rotation with the desired axis by requiring the separation to be limited to ϕ_{max}.

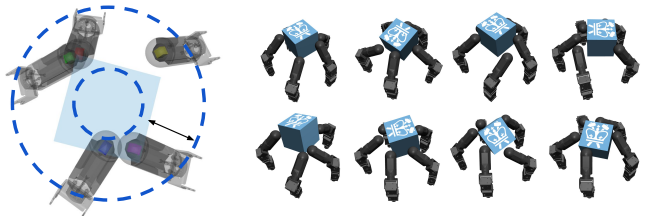
$$\begin{aligned} r = &\max(r_{max}, \boldsymbol{\omega}.\hat{\boldsymbol{k}})\ \mathbf{I}[n_c \geq 3 \wedge \phi \leq \phi_{max}] \\ &+ \min(0, \boldsymbol{\omega}.\hat{\mathbf{k}})\ \mathbf{I}[n_c < 3 \vee \phi > \phi_{max}] \end{aligned} \tag{2}$$

The reward function is described in (2), where n_c is the number of fingertip contacts and ϕ is the separation between the desired and current axis of rotation. Symbols $\wedge$, $\vee$, $\mathbf{I}$ are the logical *and*, the logical *or*, and indicator function, respectively. Notice that we also use reward clipping to avoid local optima and idiosyncratic behaviors. Although the reward uses the object's angular velocity, we do not need additional sensing to measure it as we only train in simulation with the intent of zero-shot transfer to hardware.

C. Enabling exploration with domain knowledge

A fundamental issue in using reinforcement learning for learning precision in-hand manipulation skills is that a random exploratory action can easily disturb the stability of the object held in a precision grasp, causing it to be dropped. In our case, where we are interested in learning finger-gaiting, this issue is further worsened. Finger-gaiting requires fingertips to break contact with the object and transition between different grasps, involving different fingertips, all while re-orienting the object. As one can expect, the likelihood of selecting a sequence of random actions that can accomplish this feat while obtaining a useful reward signal is extremely low.

For a policy to learn finger-gaiting, it must encounter these diverse grasps within its training samples so that the policy's action distributions can improve at these states Consider taking a sequence of random actions starting from a stable



(a) (b)

Fig. 4: (a) Sampling fingertips around the object (b) Diverse relevant initial grasps sampled for efficient exploration

m-finger grasp. While it possible to reach a stable grasp with an additional finger in contact (if available), it is more likely to lose one finger contact, then another and so on until the object is dropped. Over multiple trials, we can expect to encounter most combinations of $m-1$ grasps. In this setting, it can be argued that starting from a stable grasp with all n fingers in contact leads to maximum exploration. Interestingly, as we will demonstrate in Sec IV-A, we found this to be insufficient.

Our insight is to observe that through domain knowledge we are already aware of the states that a sufficiently exploratory policy might visit. Using domain knowledge in designing initial distributions is known way of improving sample complexity. [25][26]. Thus, we use our knowledge of relevant states in designing the initial states used for episode rollouts and show that it is critical for learning precision finger-gaiting and finger-pivoting.

We propose sampling sufficiently-varied stable grasps relevant to re-orienting the object about the desired axis and use them as initial states for collecting training trajectories. These grasps must be well distributed in terms of number of contacts, contact positions relative to the object, and object poses relevant to the task. To this end, we first initialize the object in an random pose and then sample fingertip positions until we find a stable grasp as described in Stable Grasp Sampling (SGS) in Alg. 1.

In, SGS we first sample object pose and a hand pose, then update the simulator with the sampled poses towards obtaining a grasp. We forward simulate for a short duration, t_s, to let any transients die down. If the object has settled into a grasp with at least two contacts, it is used towards collecting training trajectories. Note that the fingertips could be overlapping with the object or with each other as we do not explicitly check this. However, this is resolved during forward simulation. An illustrative set of grasps sampled by SGS are shown in Fig 4b.

Algorithm 1 Stable Grasp Sampling (SGS)

Input: ρ_{obj}, ρ_{hand}, t_s, $n_{c,\min}$ ▷ object pose distribution, hand pose distribution, simulation settling time, minimum number of contacts

Output: $\mathbf{s}_g$ ▷ simulator state of the sampled grasp

1: **repeat**
2: Sample object and hand pose: $\boldsymbol{x}_s \sim \rho_{obj}$, $\boldsymbol{q}_s \sim \rho_{hand}$
3: Set object pose in the simulator with $\boldsymbol{x}_s$
4: Set joint positions and controller set-points with $\boldsymbol{q}_s$
5: Step the simulation forward by t_s seconds
6: Find number of fingertips in contact with object, n_c
7: **until** $n_c \geq n_{c,\min}$
8: Save simulator state as $\boldsymbol{s}_g$

To sample the hand pose, we start by sampling fingertip locations within an annulus around the around the object. As shown in Fig 4a, this annulus is centered on the object and partially overlaps with it, such that probabilities of the fingertip making contact with the object and of staying free are roughly the same. With this procedure, not only do we find stable grasps relevant to finger-gaiting and finger-pivoting, we improve the likelihood of discovering them, thus minimizing training wall-clock time.

IV. Experiments and Results

For evaluating our method, we focus on learning precision in-hand re-orientation about the z- and x- axes for a range of regular object shapes. We do not separately consider y-axis re-orientation as it is similiar to x-axis, given the symmetry of our hand model. Our object set, which consists of a cylinder, sphere, icosahedron, dodecahedron and cube, is designed so that we have objects of varying difficulty with the sphere and cube being the easiest and hardest, respectively. For training, we use PPO [3].

For the following analysis we take learning z-axis re-orientation as a case study. In addition to the above, we train z-axis re-orientation policies without assuming joint set-point feedback $\mathbf{q}_d$. For all these policies, we study their robustness properties by adding noise and also by applying perturbation forces on the object (Sec IV-B). We also study the zero-shot generalization properties of these policies (Sec IV-C). Finally, through ablation studies we present a detailed analysis ascertaining the importance of different components of feedback for achieving finger-pivoting (Sec IV-D).

We note that, in simulation, the combination of $\mathbf{q}_d$ and $\mathbf{q}$ can be considered a good proxy for torque, since simulated controllers have stable stiffness. However, this feature might not transfer to a real hand, where transmissions exhibit friction, stiction and other hard to model effects. We thus evaluate our policies both with and without joint set-point observations.

A. Learning finger-gaiting manipulation

In Fig 6a, we show the learning curves for object re-orientation about the z-axis for a range of objects from using our method of sampling stable initial grasps to improve exploration. We also show learning curves using a fixed initial state (grasp with all fingers) for representative objects. First, we notice that the latter approach does not succeed. These policies only achieve small re-orientation via in-grasp manipulation and drop the object after maximum re-orientation achievable without breaking contacts.

However, when using a wide initial distribution of grasps (sampled via SGS), the policies learn finger-gaiting and

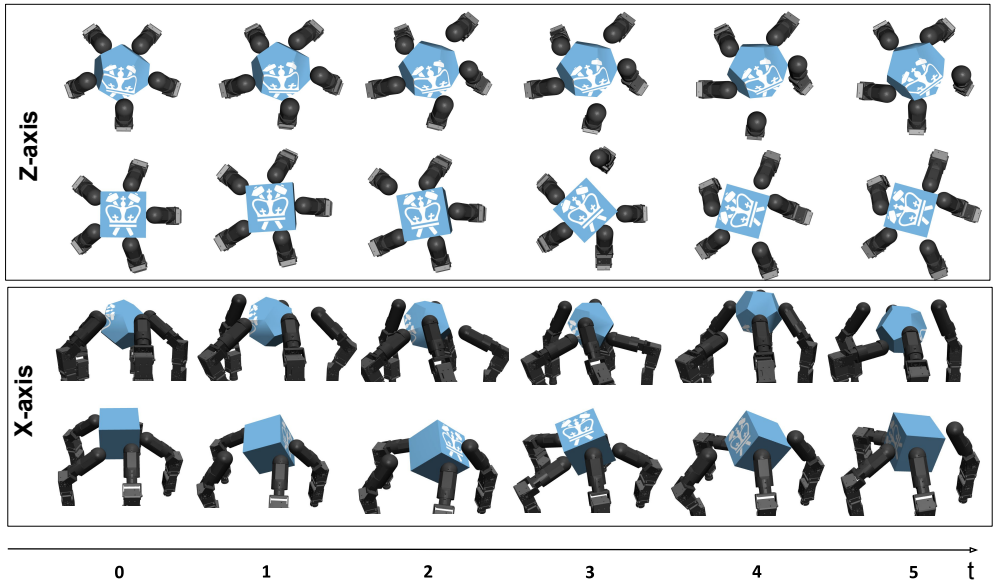


Fig. 5: Finger-gaiting and finger-pivoting our policies achieve to re-orient about z-axis and x-axis respectively. Gait frames are shown for two objects, dodecahedron and cube. Videos of the gaits can be found at https://roamlab.github.io/learnfg/

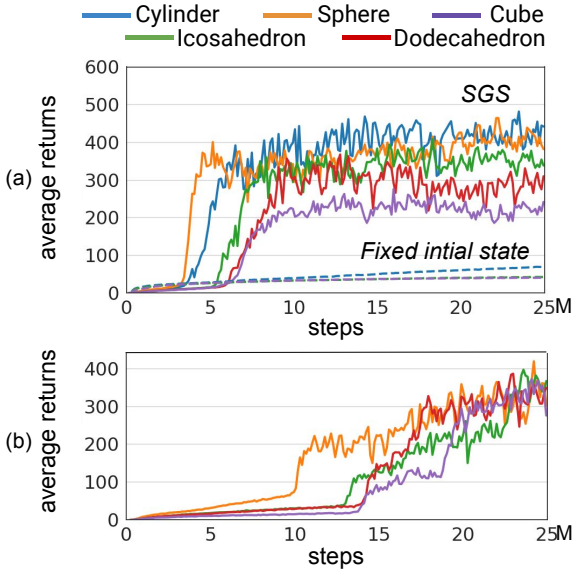


Fig. 6: Average returns for (a) z-axis re-orientation and (b) x-axis re-orientation. Figure shows that learning with wide range of initial grasps sampled via SGS succeeds while using a fixed initial state fails.

achieve continuous re-orientation of the object with significantly higher returns. With our approach, we also to learn finger-pivoting re-orientation about the x-axis with learning curves shown in Fig 6b. Thus, we empirically see that using a wide initial distribution consisting of relevant grasps is critical for learning continuous in-hand re-orientation via finger-gaiting. Fig 5 shows our finger-gaiting and finger-pivoting policies performing continuous object re-orientation about z-axis and x-axis respectively.

As expected, difficulty of rotating the objects increases as we consider objects of lower rotational symmetry from sphere to cube. In the training curves in Fig 6, we can observe this trend not only in the final returns achieved by the respective policies, but also in the number of samples required to learn continuous re-orientation. The sudden jump in the returns, which corresponds to when the policy "figures out" finger-gaiting/finger-pivoting, is also observed later for harder objects.

We also successfully learn policies for in-hand re-orientation without joint set-point position feedback, but these policies achieve slightly lower returns. However, they may have interesting consequences for generalization as we will discuss in Sec IV-C.

B. Robustness

In Fig 7, we show the performance of our policy for the most difficult object in our set (cube) as we artificially add Gaussian noise to different sensors' feedback with increasing variance. We also increasingly add perturbation forces on the object. We can see that, overall, our policies are robust to noise and perturbation forces of magnitudes reasonable for a real hand.

Our policies show little drop in performance for noise in joint positions $\mathbf{q}$. However, our policies are more sensitive to noise in contact feedback; nonetheless, they are still robust and achieve high returns even at 5mm error in contact position and 25% error in contact force. Interestingly, for noise in contact postion, we found that drop in performance arises indirectly through the error in contact normal $\hat{\mathbf{t}}_n^i$ (computed from contact position $\mathbf{c}_n^i$). As for perturbation forces on the object, we observe high returns even for at high perturbation forces (1N) equivalent to the weight of our objects. Also, the policies trained without joint-set feedback have similar robustness profiles.

C. Generalization

We study generalization properties of our policies by evaluating it on a different object in the object set. For this we consider the transfer score, which is the ratio R_{ij}/R_{ii} where R_{ij} is the average returns obtained when evaluating the policy learned with object i on object j.

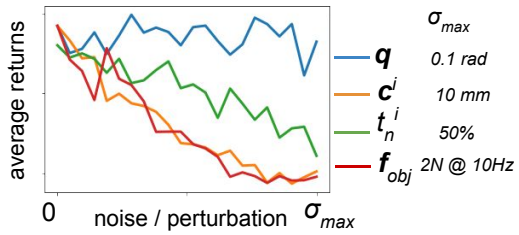


Fig. 7: Figure shows the robustness of our policies with increasing sensor noise and perturbation forces on the object.

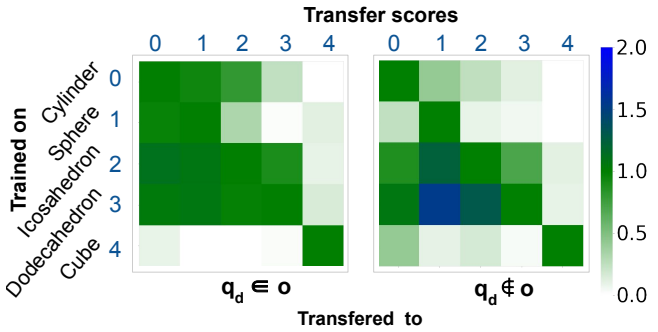


Fig. 8: Figure shows the cross transfer scores for policies with and without $\mathbf{q}_d$ in feedback

In Fig 8, we see the cross transfer performance for policies trained with all feedback. For these policies, we see that the policy trained on the sphere transfers to the cylinder and vice versa. Also, the policies trained on icosahedron and dodecahedron transfer well amongst themselves while also performing well on sphere and cylinder. Interestingly, the policy trained on the cube does not transfer well to the other objects. For policies learnt without joint set-point position feedback $\mathbf{q}_d$, the policy learned on the cube transfers to more objects. With no way to infer motor forces, the policy potentially learns to rely more on contact feedback which aids generalization.

D. Observations on feedback

While our work provides some insight w.r.t the important components of our feedback through our robustness and generalization results, it is still limited in scope. There are a number of interesting questions. Is it possible for use to learn finger-gaiting with only proprioceptive feedback? What about learning with just contact feedback? What matters in contact feedback? To answer such questions, we run a series of ablations holding out different components. For this, we again consider learning finger-gaiting on the cube as shown in Fig 9.

With this ablation study, we can make a number of important and interesting observations. As we can expect, contact feedback is essential for learning in-hand re-orientation via finger-gaiting; we find that the policy does not learn finger-gaiting with just proprioceptive feedback (#4). More interesting and also surprising is that explicitly computing contact normal $\mathbf{t}_n^i$ and providing it as feedback is critical when excluding joint position set-point $\mathbf{q}_d$ (#6 to #10). In fact, the policy learns finger-gaiting with just contact normal and joint position feedback (#10). However, while not critical,

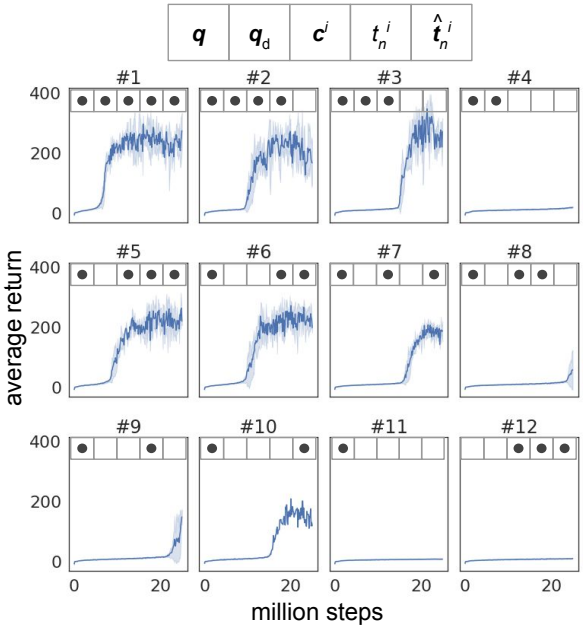


Fig. 9: Ablations holding out different components of feedback. For each experiment, dots in the observation vector shown above the training curve indicate which of the components of the observation vector are provided to the policy.

contact position and force feedback are still beneficial as they improve sample efficiency (#6, #7).

V. Conclusion

In this paper, we focus on the problem of learning in-hand manipulation policies that can achieve large angle object re-orientation via finger gaiting. To facilitate future deployment in real scenarios, we restrict ourselves to using sensing modalities intrinsic to the hand, such as touch and proprioception, with no external vision or tracking sensor providing object-specific information. Furthermore, we aim for policies that can achieve manipulation skills without using a palm or other surfaces for passive support, and which instead need to maintain the object in a stable grasp.

A critical component of our approach is the use of appropriate initial state distributions during training, used to alleviate the intrinsic instability of precision grasping. We also decompose the manipulation problem into axis-specific rotation policies in the hand coordinate frame, allowing for object-agnostic policies. Combining these, we are able to achieve the desired skills in a simulated environment, the first instance in the literature of such policies being successfully trained with intrinsic sensor data.

We consider this work to be a useful step towards future sim-to-real transfer. To this end, we engage in an exhaustive empirical analysis of the role that each sensing modality plays in enabling our manipulation skills. Specifically, we show that tactile feedback in addition to proprioceptive sensing is critical in enabling such performance. Finally, our analysis of the policies shows that they generalize to novel objects and are also sufficiently robust to force perturbations and sensing noise, suggesting the possibility of future sim-to-real transfer.

References

[1] A M Okamura, N Smaby, and M R Cutkosky. "An overview of dexterous manipulation". In: *Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No.00CH37065)*. Vol. 1. Apr. 2000, 255–262 vol.1.

[2] Y Tassa, T Erez, and E Todorov. "Synthesis and stabilization of complex behaviors through online trajectory optimization". In: *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. Oct. 2012, pp. 4906–4913.

[3] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. "Proximal Policy Optimization Algorithms". In: July 2017. arXiv: 1707.06347 [cs.LG].

[4] OpenAI, Marcin Andrychowicz, Bowen Baker, Maciek Chociej, Rafal Jozefowicz, Bob McGrew, Jakub Pachocki, Arthur Petron, Matthias Plappert, Glenn Powell, Alex Ray, Jonas Schneider, Szymon Sidor, Josh Tobin, Peter Welinder, Lilian Weng, and Wojciech Zaremba. "Learning Dexterous In-Hand Manipulation". In: (Aug. 2018). arXiv: 1808.00177 [cs.LG].

[5] Qiang Li, Oliver Kroemer, Zhe Su, Filipe Fernandes Veiga, Mohsen Kaboli, and Helge Joachim Ritter. "A Review of Tactile Information: Perception and Action Through Touch". In: *IEEE Trans. Rob.* 36.6 (Dec. 2020), pp. 1619–1634.

[6] P Michelman. "Precision object manipulation with a multifingered robot hand". In: *IEEE Trans. Rob. Autom.* 14.1 (Feb. 1998), pp. 105–113.

[7] Susanna Leveroni and Kenneth Salisbury. "Reorienting Objects with a Robot Hand Using Grasp Gaits". In: *Robotics Research*. Springer London, 1996, pp. 39–51.

[8] L Han and J C Trinkle. "Dextrous manipulation by rolling and finger gaiting". In: *Proceedings. 1998 IEEE International Conference on Robotics and Automation (Cat. No.98CH36146)*. Vol. 1. May 1998, 730–735 vol.1.

[9] R Platt, A H Fagg, and R A Grupen. "Manipulation gaits: sequences of grasp control tasks". In: *IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA '04. 2004*. Vol. 1. Apr. 2004, 801–806 Vol.1.

[10] Jean-Philippe Saut, Anis Sahbani, Sahar El-Khoury, and Veronique Perdereau. "Dexterous manipulation planning using probabilistic roadmaps in continuous grasp subspaces". In: *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*. Oct. 2007, pp. 2907–2912.

[11] T Omata and M A Farooqi. "Regrasps by a multi-fingered hand based on primitives". In: *Proceedings of IEEE International Conference on Robotics and Automation*. Vol. 3. Apr. 1996, 2774–2780 vol.3.

[12] Yongxiang Fan, Te Tang, Hsien-Chung Lin, Yu Zhao, and Masayoshi Tomizuka. "Real-Time Robust Finger Gaits Planning under Object Shape and Dynamics Uncertainties". In: (Oct. 2017). arXiv: 1710.10350 [cs.RO].

[13] Balakumar Sundaralingam and Tucker Hermans. "Geometric In-Hand Regrasp Planning: Alternating Optimization of Finger Gaits and In-Grasp Manipulation". In: (Apr. 2018). arXiv: 1804.04292 [cs.RO].

[14] Tingguang Li, Krishnan Srinivasan, Max Qing-Hu Meng, Wenzhen Yuan, and Jeannette Bohg. "Learning Hierarchical Control for Robust In-Hand Manipulation". In: (Oct. 2019). arXiv: 1910.10985 [cs.RO].

[15] Filipe Veiga, Riad Akrour, Jan Peters, and. "Hierarchical Tactile-Based Control Decomposition of Dexterous In-Hand Manipulation Tasks". en. In: *Front Robot AI* 7 (Nov. 2020), p. 521448.

[16] Fan Shi, Timon Homberger, Joonho Lee, Takahiro Miki, Moju Zhao, Farbod Farshidian, Kei Okada, Masayuki Inaba, and Marco Hutter. "Circus ANYmal: A Quadruped Learning Dexterous Manipulation with Its Limbs". In: (Nov. 2020). arXiv: 2011.08811 [cs.RO].

[17] Andrew S Morgan, Daljeet Nandha, Georgia Chalvatzaki, Carlo D'Eramo, Aaron M Dollar, and Jan Peters. "Model Predictive Actor-Critic: Accelerating Robot Skill Acquisition with Deep Reinforcement Learning". In: (Mar. 2021). arXiv: 2103.13842 [cs.RO].

[18] Aravind Rajeswaran, Vikash Kumar, Abhishek Gupta, Giulia Vezzani, John Schulman, Emanuel Todorov, and Sergey Levine. "Learning Complex Dexterous Manipulation with Deep Reinforcement Learning and Demonstrations". In: (Sept. 2017). arXiv: 1709.10087 [cs.LG].

[19] Henry Zhu, Abhishek Gupta, Aravind Rajeswaran, Sergey Levine, and Vikash Kumar. "Dexterous Manipulation with Deep Reinforcement Learning: Efficient, General, and Low-Cost". In: *CoRR* abs/1810.06045 (2018). arXiv: 1810.06045. URL: http://arxiv.org/abs/1810.06045.

[20] Ilija Radosavovic, Xiaolong Wang, Lerrel Pinto, and Jitendra Malik. "State-Only Imitation Learning for Dexterous Manipulation". In: (Apr. 2020). arXiv: 2004.04650 [cs.RO].

[21] A Nagabandi, K Konolige, S Levine, et al. “Deep dynamics models for learning dexterous manipulation”. In: *Conference on Robot* (2020).

[22] Herke van Hoof, Tucker Hermans, Gerhard Neumann, and Jan Peters. “Learning robot in-hand manipulation with tactile features”. In: *2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids)*. Nov. 2015, pp. 121–127.

[23] Andrew Melnik, Luca Lach, Matthias Plappert, Timo Korthals, Robert Haschke, and Helge Ritter. “Using Tactile Sensing to Improve the Sample Efficiency and Performance of Deep Deterministic Policy Gradients for Simulated In-Hand Manipulation Tasks”. en. In: *Front Robot AI* 8 (June 2021), p. 538773.

[24] Tao Chen, Jie Xu, and Pulkit Agrawal. “A Simple Method for Complex In-hand Manipulation”. In: *5th Annual Conference on Robot Learning*. 2021. URL: https://openreview.net/forum?id=7uSBJDoP7tY.

[25] Sham Machandranath Kakade. “On the sample complexity of reinforcement learning”. en. PhD thesis. Ann Arbor, United States: University of London, University College London (United Kingdom), 2003.

[26] D P de Farias and B Van Roy. “The linear programming approach to approximate dynamic programming”. In: *Oper. Res.* 51.6 (Dec. 2003), pp. 850–865.