Haptic-based and $SE(3)$-aware Object Insertion using Compliant Hands

Osher Azulay, Maxim Monastirsky and Avishai Sintov

***Abstract*—Object insertion is primarily studied using rigid robotic hands. However, these may have difficulties overcoming spatial uncertainties originating from an uncertain initial grasp. Compliant hands, on the other hand, can cope with $SE(3)$ uncertainties and adapt to the environment upon contact. Nevertheless, contact forces may contribute additional uncertainties and lead to failure if not controlled properly. In this letter, we take inspiration from human insertion and study how *haptic glances* with compliant hands during contact can provide valuable information regarding object state. Using a force/torque sensor, we show that a haptic glance based on excitation of finger perturbations can provide accurate contact localization and indication of a successful insertion. With such insight, we propose an online learning scheme for general precision control of contact-rich object insertion. A deep residual Reinforcement Learning (RL) policy leverages the contact dynamics of the compliant hand to cope with $SE(3)$ uncertainties. Several experiments of precision insertion tasks with various objects and grasp uncertainties exhibit high success rate and validate the effectiveness of the approach.**

***Index Terms*—Kinesthetic, pose estimation, haptic glance.**

I. Introduction

DAILY life activities require interaction with the environment through contact. For a robot to perform real-world tasks, it must be able to work in unstructured contact-rich environments. *Insertion* is a common problem that requires dealing with contacts and has been extensively researched [1]. The insertion, or *peg-in-a-hole*, operation occurs in various tasks such as assembly [2] and dense packing [3]. While humans can perform insertion assembly with relative ease, achieving high-precision and robust insertion in unstructured environments remains a challenge for robotic systems.

The vast majority of work on robot insertion has been carried out by rigid hands. Rigid hands, however, may have difficulty coping with uncertainties originating from erroneous object pose estimation and environment perception. In particular, accurate relative poses of hand-object and object-hole are usually not available in real-world applications. Therefore, compliance of the robotic arm has been incorporated to enable it to mechanically adapt to the environment upon contact while complicating the hardware [5]. Contrary to artificial compliance solutions, an *Underactuated Compliant Hand* (UCH), seen in Figure 1, is a grasping mechanism that inherently incorporates compliance [6], [7]. Consequently, they have gained popularity in recent years due to their low-cost and ability to maintain a stable grasp with open-loop control. In addition, UCH have been shown efficient for in-hand and precise manipulation [8], [9]. However and due to their compliant nature, large pose uncertainties of the object with regards to the hand exist [10], [11]. Consequently, visual perception is the leading approach to acquire pose estimation of the object between the fingers [12]–[14]. Hence, the sole use of UCH for insertion tasks was based on an RGB-Depth camera [15]. The work has proposed a vision-driven compliant method using depth-based pose estimation to control planar insertion with a UCH.

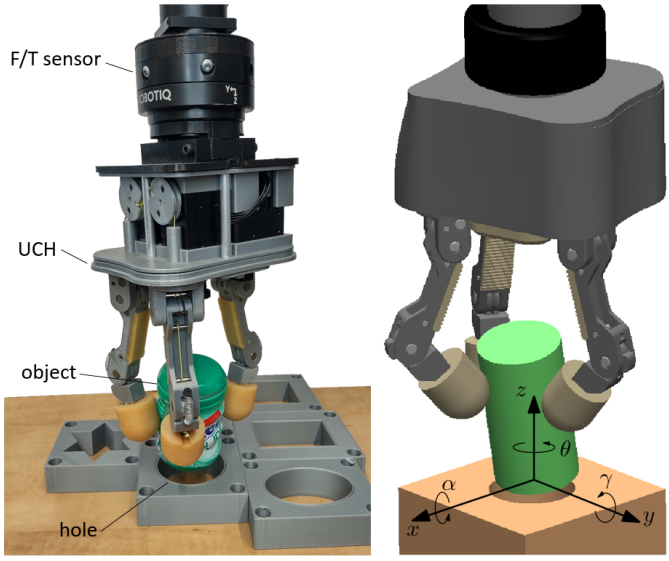


Fig. 1. A Model-O [4] Underactuated Compliant Hand (UCH) inserting an object into a hole using haptic glances from an F/T sensor. Due to the compliance of the hand, the method is able to consider $SE(3)$ pose uncertainties of the object including roll α and pitch γ tilt angles.

A complete insertion solution solely based on visual perception may be limited in an unstructured or cluttered environments. In addition, visual perception might provide limited accuracy for a tight-tolerance insertion problem [16]. Therefore, extensive haptic work has focused on kinesthetic- or tactile-based exploration at the final insertion phase given an approximated position of the hole acquired from a camera. The common use of kinesthetic haptics is based on a Force/Torque (F/T) sensor mounted between a rigid hand and arm (i.e., wrist) [16], [17]. In tactile-based insertion, touch sensors on the fingertips of the hand provide embedded information

Manuscript received: September, 13, 2022; Revised October, 30, 2022; Accepted November, 20, 2022.

This paper was recommended for publication by Editor Ashis Banerjee upon evaluation of the Associate Editor and Reviewers' comments. This work was supported by the Pazy Foundation (grant No. 283-20).

O. Azulay, M. Monastirsky and A. Sintov are with the School of Mechanical Engineering, Tel-Aviv University, Israel. e-mail: {osherazulay,maximm}@mail.tau.ac.il, sintov1@tauex.tau.ac.il

Digital Object Identifier (DOI): see top of this page.

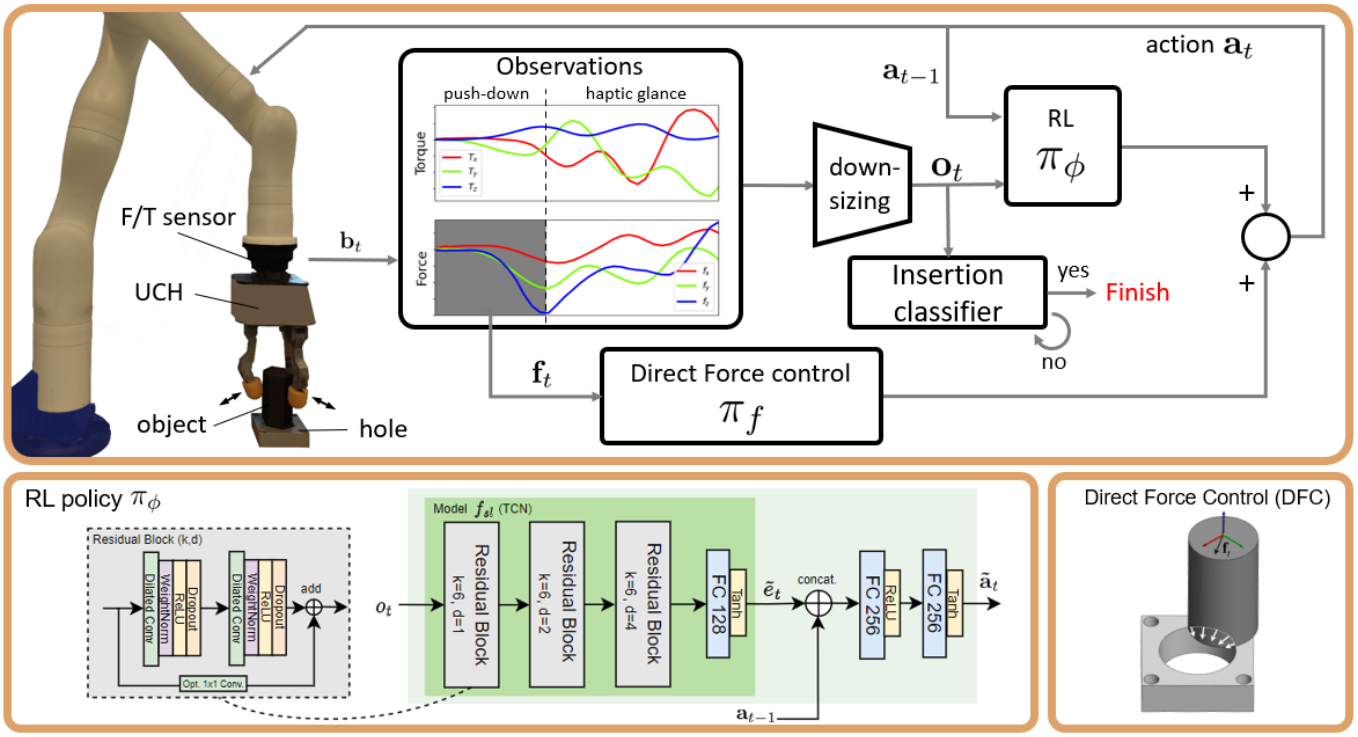


Fig. 2. Illustration (top) of the proposed workflow for insertion with an Underactuated Compliant Hand (UCH). An observation is acquired through a haptic glance where a set of F/T measurements are taken during initiated contact with the hole. Simple force control (bottom right) is used to expedite the training of a Reinforcement Learning (RL) policy (bottom left) to perform the task. An insertion classifier checks, based on the haptic glances, whether the object was successfully inserted.

regarding hand-object and object-environment contacts [18], [19]. While tactile sensors are required to be integrated and sometimes customized to the hand itself, standard F/T sensors are off-the-shelve devices frequently present on robotic arms and independent of the hand type. The majority of prior work on insertion considered either only the position of the object or including also just the planar angular alignment (i.e., yaw angle) with the hole. In particular, no work has considered the entire $SE(3)$ pose of the object with respect to the hole using only F/T sensing.

State-of-the-art in haptic-based insertion is all based on rigid hands. Hence, the insertion problem by UCH solely based on haptics has never been addressed. For UCH, however, integrating tactile sensors would make it harder to modulate and replace fingers for different tasks in low-cost open-source hardware such as in [4] if tactile sensors are to be integrated. On the other hand, mounting an F/T sensor between the arm and hand is an easy task. Compared to vision, F/T sensing provides also implicit information regarding object-environment surface quality. However and due to compliance, the system would be much more sensitive to object perturbation due to contact with the environment than rigid hands. Consequently, load attenuation of the springs along with object pose perturbations may complicate F/T signal processing. Nevertheless and as demonstrated in previous work [15], [20], compliance provides assistance in insertion and motivates the use of UCH. UCH which was demonstrated to be efficient in various tasks has yet to be explored for insertion while including: spatial position and orientation deviations of the object relative to the hole, and F/T-based sensing.

In this letter, we tackle the problem of F/T-based insertion with UCH while observing the ability of the hand to correct $SE(3)$ deviations. Rather than having a soft or compliant robot arm, which can be of high cost, a low cost compliant hand is easier to be used on any robotic arm. First, we explore the notion of an *haptic glance* in such system. A haptic glance is a robot initiated object contact in the vicinity of the hole in order to gain information regarding the instantaneous pose (i.e., position and orientation) of the object relative to the hole. We show that finger perturbations during contact with the hole is the most effective and safe haptic glance. The haptic glance is also shown to be able to classify whether the object was successfully inserted. Next, we use haptic glances to propose a data-efficient approach to train a Reinforcement Learning (RL) policy for insertion (Figure 2). The training involves a Direct Force Controller (DFC) to guide and expedite the learning. Curriculum learning is also utilized for training from easier to more complex objects.

To the best of the author's knowledge, this is the first attempt to perform robotic insertion with a UCH solely based on haptic F/T sensing. In addition and contrary to previous work with F/T sensing, this work considers $SE(3)$ deviations of the object with respect to the hole. Hence, the contribution of this work is novel RL-based framework to cope with $SE(3)$ uncertainties and align the object into the hole. The framework includes a haptic glance that utilizes the ability of the hand for contact localization and insertion classification. A simple but effective controller is used to reduce data sampling. The method can generalize to novel objects of different shapes and hole clearances.

TABLE I
STATE-OF-THE-ART COMPARISON FOR ROBOTIC INSERTION

Work		RL	Vision	F/T	Tactile	R\C[1]
Van-Wyk et al.	[27]	✗	✗	✓	✓	R
Lee et al.	[25]	✓	✓	✓	✗	R
Ding et al.	[24]	✓	✗	✓	✗	R
Beltran et al.	[16]	✓	✗	✓	✗	R
Morgan et al.	[15]	✗	✓	✗	✗	C
Cui et al.	[28]	✗	✗	✗	✓	R
Kim et al.	[19]	✓	✗	✗	✓	R
Ours		✓	✗	✓	✗	C

[1]R\C denotes Rigid or Compliant hands.

II. Related work

Several early studies suggested the use of passive hardware-based compliance, where a mechanical device, termed Remote Center Compliance (RCC), is positioned between the arm and hand to aid the insertion [20]. By virtue of its passive compliance, the RCC enables the gripper to move perpendicular to the peg's axis and rotate freely to reduce resistance. However, RCC is less effective with high-precision assembly [21]. On the other hand, active compliant methods, where some force control is employed, are widely used to overcome uncertainty [22], [23].

Table II provides a comparative summary of state-of-the-art work on insertion methods. Recent work used Deep Reinforcement Learning (DRL) along with an F/T sensor and active compliance to drive a peg into a hole [16]. Similarly, an F/T sensor was used to insert an object held by a parallel gripper based on a learned dynamics model [24]. Another work fused F/T sensing and RGB images to improve sample efficiency in policy learning [25]. The latter two, however, focused only on positional localization and did not consider angular deviations with respect to the hole.

Contrary to F/T-based insertion which is extrinsic to the hand, tactile sensors are embedded within the fingers. Optic-based sensors were embedded on a parallel gripper to estimate contact location in insertion of small parts [18]. Similar sensors were later used in tight box-packing [26] and general insertion [19]. A different approach equipped a multi-finger hand with F/T sensors at the tip of the fingers [27]. As discussed above, tactile sensors are less appealing to UCH. Therefore, the use of an extrinsic F/T sensor with UCH is explored in this work.

III. Method

A. Problem formulation

We consider a robotic arm equipped with an F/T sensor and a three-fingered UCH. While the UCH is grasping the object, as seen in Figure 2, the system will attempt to insert the object into a near-by hole under pose uncertainties. The hole is in the shape of the object with some clearance ϵ. The state of the object $\mathbf{s} \in SE(3)$ includes its position (x, y, z) and orientation (α, γ, θ) relative to the hole (Figure 1). Prior work with rigid hands and F/T sensing only considered the $SE(2)$ of the object-hole (i.e., position (x, y) and yaw angle θ). In UCH, however, the roll α and pitch γ can be uncertain due to the initial grasp or involuntary changes originating from contact forces and hand compliance. Explicit access to the true state is not available. The system can only acquire F/T measurements $\mathbf{b} \in \mathcal{B}$ for $\mathcal{B} \subset \mathbb{R}^6$. Hence, an observation $\mathbf{o}_t \in \mathcal{O}$ at time t, where $\mathcal{O} \subset \mathcal{B} \times \ldots \times \mathcal{B}$, is a set of F/T measurements during an initiated temporal contact.

While correction of x, y and θ in $\mathbf{s}$ requires planar actions of the hand by the arm, correction of α and γ requires either spatial rotation of the arm or finger manipulation. The former may not be possible in a cluttered environment and the latter requires planned in-hand manipulation which is not in the scope of this work. Nevertheless, it is hypothesized that learning planar motion of the object on the hole surface could utilize compliance to correct object roll and pitch tilts. Hence, let $\mathbf{a} \in \mathcal{A}$ for $\mathcal{A} \subset \mathbb{R}^3$ be a continuous robot action displacement. Action $\mathbf{a} = (\Delta x, \Delta y, \Delta\theta)^T$ includes the required planar position $(\Delta x, \Delta y)$ and yaw angle $\Delta\theta$ changes. The insertion goal is, therefore, to align the object and insert it into the hole under the following assumptions: No prior knowledge of the object shape nor the target hole are available. Yet, the shapes of the object and target hole are assumed to match; State-of-the-art vision system can roughly estimate the pose of the hole up to few millimeters and degrees tolerance [16]; The object was previously grasped with uncertainty regarding its $SE(3)$ state; The final precise insertion would solely be based on F/T feedback. Due to the above assumptions, the target location of the hand-arm system for object-hole alignment is not known.

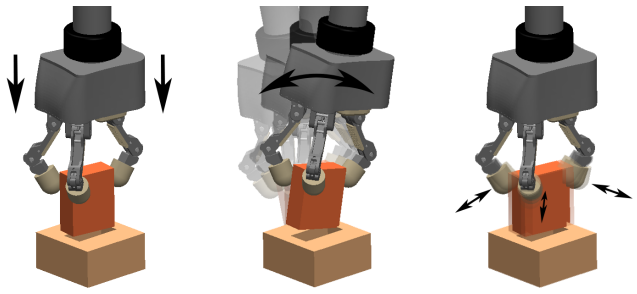
B. Haptic Glance for contact localization

Since the true state $\mathbf{s}$ of the object is not known, the robot makes an observation $\mathbf{o}$. An observation is, in fact, an *haptic glance* [29] where contact is initiated with the environment in order to either explicitly approximate the state or evaluate the residual $\mathbf{a}$ with respect to the hole towards alignment with it. The core modality in a haptic glance is F/T sensing.

Taking inspiration from human motion during insertion, we have investigated three multi-phased haptic glance primitives. The primitives are based on either arm or hand motion. Nevertheless, they all start from a *push-down* phase of lowering down the grasped object towards the surface of the hole until reaching a normal force threshold indicating contact. A single F/T measurement is ambiguous for accurate state estimation. Hence and once in contact, the haptic exploration phase can be conducted across a certain time frame while collecting diverse F/T data. We investigate the use of the following three exploration primitives.

Vertical press (Figure 3a). The baseline exploration motion would be to keep pushing down the object towards the surface of the hole while increasing and decreasing the force exerted by the arm [30]. Due to intersection with the hole along with the hand compliance, the object would slightly tilt accordingly and may provide informative F/T signals.

Arm tilts (Figure 3b). Inspired by [19], [24], object tilting is deliberately exerted by the arm in order to sense the edge of the hole. Hence, the arm is tilted back and forth to three directions with respect to the surface normal. The center of each tilt is approximately the contact point with the hole. While we



(a) (b) (c)

Fig. 3. Three types of haptic glances: (a) vertical press, (b) arm tilts and (c) finger perturbations.

acknowledge that arm tilts may not be possible in cluttered environments, we wish to investigate performance.

Finger perturbations (Figure 3c). While the previous haptic glance primitives were based on arm movements, here we exploit the fingers ability to press against the object. Each finger, in its turn, would gradually increase contact force with the object. Due to finger compliance, the object would slightly move and tilt based on its position relative to the hole and provide corresponding F/T measurements.

Tilting the entire arm in confined spaces may not be possible due to the risk of collision. Moreover, it is hypothesized that the finger perturbations approach would be safer from object jamming than arm tilts. In addition, finger perturbations slightly move the object arbitrary on the plane with some probability that it may align with the hole. When aligned and due to compliance, the object would passively slide inside. Hence, this primitive is expected to have a higher success rate. Consequently, the use of only the fingers to press on the object is safer, easier and more appealing. It is important to note that all three motion primitives passively exploit hand compliance. During the haptic glance and due to the initial load phase, the hand acts as a pressed spring. Hence, the hand is more sensitive to the explorative motions described above which are captured by the F/T sensor.

C. Haptic Glance Encoding

The above haptic glances provide contact-rich data across some constant time frame. In particular, a haptic glance starting at time t is a temporal set of $N+1$ F/T measurements $\mathbf{o}_t = \mathbf{b}_{[t:t+N]}$. Recording with high-frequency would provide rich temporal data while may also contain repetitions. Hence, training a model with the entire data in $\mathbf{o}_t$ may be redundant. Therefore, we encode the temporal data by down-sizing [26]. The down-sizing is done by averaging K consecutive bins along the haptic glance set. Consequently, the encoding now yields a smaller set $\mathbf{o}_t \in \mathcal{O}$ of K samples. Haptic glance $\mathbf{o}_t$ is now the new observation for contact localization. The same encoding applies for all three exploration primitives.

D. Insertion success identification

Prior to training and testing insertion, the system must first identify successful insertions. As opposed to rigid hands and due to compliance, an object grasped by a UCH can slide into the hole without lowering the hand by the arm. Hence, it is required to identify successful insertion during and in between haptic glances. For such, we collect $2M$ haptic glances $\{\mathbf{o}_i\}_{i=1}^{2M}$ where M glances are when the object is inserted and M glances are in contact with the surface of the hole. Each glance $\mathbf{o}_i$, is also labeled with $\mathbf{d}_i = \{\texttt{in}, \texttt{out}\}$ indicating whether the object is in the hole or not. This is used to train a Temporal Convolutional Network (TCN) [31] classifier that provides the probability of whether the object was successfully inserted and the episode can be finished.

E. Direct Force Control

We propose a naive force control to correct the position of the object towards the hole. The Direct Force Control (DFC) is proposed as a position insertion baseline and used also for augmenting the DRL as to be discussed below. F/T measurements, while pushing down the object on the hole, can provide implicit information regarding the planer direction towards to the hole. In other words, during the push-down phase, the object would slightly tilt approximately towards the center of the hole. Observing only the measured force, such tilt would be sensed by the F/T sensor as a tangential (i.e., in the $x-y$ plane) reaction force opposite to the tilting direction. Hence, the position can be corrected in direction opposite to the measured force.

In this control, we consider only the observation at the time frame of the push-down phase. As single force measurement can be ambiguous, we take the mean $\mathbf{f}_t \in \mathbb{R}^2$ of the measured forces from haptic glance $\mathbf{o}_t$ within the push-down phase. In other words, force $\mathbf{f}_t$ is the mean of the measured forces projected onto the $x-y$ plane (Figure 2). Assuming point contact, static motion and some object overlapping with the hole, an action proportional to $\mathbf{f}_t$ would drive the object towards the hole. Therefore and given an observation $\mathbf{f}_t$, a naive planar insertion controller π_f is of the form

$$\pi_f(\mathbf{f}_t) = -K_p \mathbf{R}_z \mathbf{f}_t. \tag{1}$$

where $K_p > 0$ is a diagonal gain matrix. Matrix $\mathbf{R}_z(\Delta\theta) \in SO(2)$ is the z-axis rotation correction if some orientation action $\Delta\theta$ was performed after the observation. The position control action is, therefore, $\tilde{\mathbf{a}}_t = (\Delta x, \Delta y)^T = \pi_f(\mathbf{f}_t)$. It is noted that F/T measurements in tilting may be ambiguous due to complex edge shapes. In addition, the DFC does not consider nor try to correct orientation misalignment. Hence, if the object is not close to alignment with the hole, we expect a low insertion success rate. Therefore, we next discuss SL and RL approaches for insertion. Nevertheless, we use the DFC to make RL more sample efficient.

F. Supervised learning model

A haptic glance includes an implicit and complex spatiotemporal signature that is difficult to model analytically [1]. Therefore, we begin by training an SL model f_{sl} to directly estimate the contact error $\mathbf{e}_t \in \mathbb{R}^3$ given a haptic glance $\mathbf{o}_t$, i.e., $\tilde{\mathbf{e}}_t = f_{sl}(\mathbf{o}_t)$ where $\tilde{\mathbf{e}}_t$ is the approximation of $\mathbf{e}_t$. The approach is inspired by object insertion with a parallel gripper and tactile sensing proposed in [26]. The SL model is trained by collecting Z pairs of observations and labels $\{\mathbf{o}_i, \mathbf{e}_i\}_{i=1}^{Z}$.

During inference upon a haptic glance, an SL-based action is given by

$$\mathbf{a}_t = -\tilde{\mathbf{e}}_t = -f_{sl}(\mathbf{o}_t). \tag{2}$$

Since the observation is comprised of temporal data, we use the TCN architecture for regression of sequential data. The TCN model is used to implement f_{sl} and predict the state error in real-time. However, the collection of data to train the SL model is done randomly and may be inefficient. RL, on the other hand, can provide a sample efficient policy by directing the learning along with optimizing the required haptic glances. In addition, the SL model cannot correct involuntary deviations in tilt angles α and γ. Unsupervised RL, discussed next, can implicitly observe these deviations from F/T signals and act to correct them.

G. Deep reinforcement learning policy

As discussed above, explicit state information is not available. Therefore, the system is assumed to be a Partial Observed Markov Decision Process (POMDP). The POMDP can be formally defined by the tuple $(\mathcal{S}, \mathcal{A}, \mathcal{O}, \mathcal{P}, \mathcal{Z}, \mathcal{R}, \delta)$ where $\mathcal{P}$, $\mathcal{Z}$ and δ are the set of conditional transition probabilities, the set of observation probabilities and discount factor, respectively. Map $\mathcal{R} : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ is the reward function. In POMDP, The next state $\mathbf{s}_{t+1} \in \mathcal{S}$ depends solely on the current state $\mathbf{s}_t \in \mathcal{S}$ and desired action $\mathbf{a}_t \in \mathcal{A}$ according to latent environment transition dynamics $\mathcal{P}(\mathbf{s}_{t+1}|\mathbf{s}_t, \mathbf{a}_t)$ while receiving a reward $r_t \in \mathcal{R}$. While the true state is not available, an observation is provided to the agent in the form of a haptic glance from $\mathcal{O}$. Therefore, the agent takes a haptic glance $\mathbf{o}_{t+1} \in \mathcal{O}$ when reaching the next state $\mathbf{s}_{t+1}$ with probability $\mathcal{Z}(\mathbf{o}_{t+1}|\mathbf{s}_{t+1})$. RL would, therefore, aim to learn a policy $\pi_\phi(\mathbf{a}_t|\mathbf{o}_t)$ that maximizes the expected return $\mathbb{E}\{\sum_{t=0}^{H} \delta^t r_t\}$, where ϕ is the vector of model parameters and H is the horizon. By modeling the system as a POMDP and using RL, the agent has the potential to estimate the true state error and align the object within a sequence of actions rather than applying greedy actions, such as with the SL in Section III-F.

An episode of the RL framework consists of one or more haptic glances. At each haptic glance $\mathbf{o}_t$, the robot *explores* the current state. In addition, the policy also includes the previous action $\mathbf{a}_{t-1}$ along with the observation $\mathbf{o}_t$ to provide implicit information about the contact configuration. Consequently, policy $\pi_\phi(\mathbf{o}_t, \mathbf{a}_{t-1})$ is trained, through a set of episodes, to apply actions that would align and insert the object to the hole with a minimal number of haptic glances. Therefore, the reward

$$r_t = -|\mathbf{e_t}| + r_i \tag{3}$$

at time t penalizes for large planar alignment errors $|\mathbf{e_t}|$ and rewards for successful insertion such that r_i equals either $\lambda > 0$ or 0 for successful and failed insertion, respectively. Instead of F/T observation $\mathbf{o}_t$, during training an asymmetric input is used in the critic model, directly feeding the true contact error $\mathbf{e}_t$ and the action from the actor model. It has been shown that using asymmetric input can significantly improve performance on real robot learning [32].

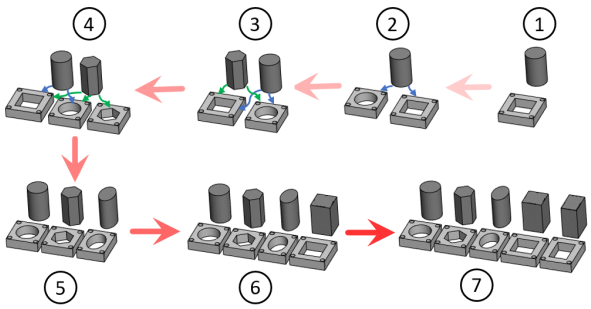


Fig. 4. Seven Curriculum Learning (CL) levels for five training objects.

H. Deep Residual Reinforcement Learning

The learning process for a real robot may be hard, time consuming and requires safe exploration. In particular, learning policy π_ϕ with no prior information may require a large amount of samples. In order to expedite the learning and provide some sample efficiency, it is proposed to bootstrap the learning of the positional alignment with the DRC policy. Hence, the position insertion action consists of two components. First, the DRC policy π_f provides a rough approximation of the positional correction. Then, RL with reward (3) learns implicit residual correction of the DRC. In other words, the DRC policy π_f provides a prior to the RL policy π_ϕ. The final Residual RL (ResRL) control policy is, therefore, a superposition of both control signals and is given by

$$\mathbf{a}_t = \pi_f(\mathbf{f}_t) + \pi_\phi(\mathbf{o}_t, \mathbf{a}_{t-1}). \tag{4}$$

While policy π_f is a naive approximation for positional alignment, policy π_ϕ observes its behaviour and learns to correct it. Note that the DFC only outputs positional action $\tilde{\mathbf{a}}_t = (\Delta x, \Delta y)$. Thus, action (4) is related only to position control while orientation action is chosen by $\pi_\phi(\mathbf{o_t}, \mathbf{a}_{t-1})$.

I. Curriculum Learning

Since sample efficiency is a key factor for real robot learning, a Curriculum-based Learning (CL) approach is also explored to further accelerate the the training. In the absence of a curriculum, the robot simply selects random object-hole pairs from episode to episode. In a curriculum strategy, the training is introduced with data in an organized order from easy samples to hard ones. In other words, more difficult aspects of a problem are gradually introduced so that the model is always challenged.

Training of the RL policies includes five primitive objects for generalization: cylinder, cuboid, square prism, hexagonal prism and elliptical cylinder. Learning to insert a cylinder or an hexagonal prism, for instance, is much simpler than a cuboid since they are less sensitive to rotational misalignment [30]. In addition, learning to insert a cylinder into a square hole is easier then to a cylindrical one. Hence, a CL strategy is designed for seven levels as illustrated in Figure 4, starting from an easier level and scaling-up the complexity. Once a success rate threshold is reached for the current level, the robot advances to the next level.

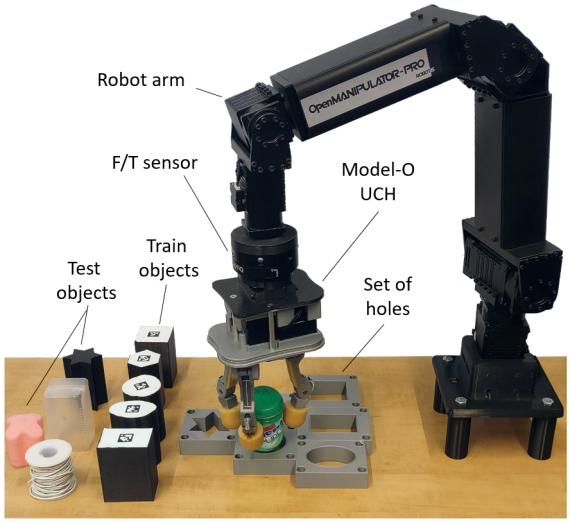


Fig. 5. The experimental setup consisting of an OpenMANIPULATOR-P robot arm, a Model-O UCH, the FT-300 F/T sensor, five training objects and their corresponding holes, and five testing objects (including the gum box within the hand).

IV. Experiments

A. Setup

Robotic hardware. An experimental setup was built to automate the training and testing of the insertion. The setup, seen in Figure 5, consists the three-finger Model-O UCH [4]. Each finger of the UCH has two compliant joints with springs where a tendon wire runs along its length and is connected to an actuator. The UCH is mounted to an OpenMANIPULATOR-P arm equipped with a Robotiq FT-300 F/T sensor. The system is operated by the Robot Operating System (ROS). During the experiment, data stream from the F/T sensor is available in 100 Hz. Velocity of the arm during insertion attempts is set manually such that inertial forces can be neglected. The motion velocity is similar for all objects and experiments.

Objects and holes. Five 3D printed objects are used for training including a cylinder, a cuboid, a square prism, an hexagonal prism and an elliptical cylinder. A matching hole is paired to each object with a clearance of $\epsilon = 1.5mm$. An additional five novel objects not included in the training are used for testing the policies and include a gum box, wire roll, hollow box, star prism and a flower shaped prism. The flower shape prism is from a shape sorting kids toy and the robot must insert it to a matching hole on a puzzle board. Other test objects are fitted into the existing holes of the training objects and some have tighter clearance. Only for ground-truth analysis of object inclinations, a camera mounted at the base of the UCH observes fiducial markers positioned on top of the objects.

Uncertainties. It is assumed that, in a real-world scenario, visual perception would provide a noisy pose estimation of the hole. To simulate this, the pose error of the object is randomly sampled from a uniform distribution. In particular, the translation and orientation errors are uniformly sampled from the range $x, y \sim [-15\ mm, 15mm]$ and $\theta \sim [-12°, 12°]$, respectively. State-of-the-art in object insertion reported up to similar values of initial errors [16], [30].

SL model. To implement f_{sl} by SL, a TCN model is trained for each primitive. Hyper-parameters of the model were optimized to the following architecture. We adopted the TCN implementation from [31] seen in Figure 2 (bottom left) with three residual blocks of 128 channels each. The TCN takes as input the observation $\mathbf{o}_t$ and processes it through a series of residual blocks, following by a Fully-Connected (FC) layer and Tanh activation. The Adam optimizer was used along with the mean squared error loss function.

TABLE II
Performance of SL polices

Primitive	Object	Cyl.	Hex. prism	Ellip. cylinder	Square prism	Cuboid
Vertical press [30]	Mean Err.	Pos. 4.0±0.6 (mm); Orien. 3.8°±0.9°				
	Success	87%	70%	33%	23%	16%
	# Glances	4.1	4.0	2.8	2.2	2.6
Arm tilts [19], [24]	Mean Err.	Pos. 3.1±0.5 (mm); Orien. 3.2°±0.6°				
	Success	80%	87%	90%	77%	63%
	# Glances	3.5	4.7	3.8	5.7	4.8
Finger perturb.	Mean Err.	Pos. 2.7±0.5 (mm); Orien. 3.3°±0.5°				
	Success	93%	93%	93%	83%	67%
	# Glances	4.4	5.3	4.9	4.3	5.3

Insertion classifier. To implement the insertion classifier, the same TCN architecture described above was used followed by a Sigmoid output instead of the Tanh activation. The classifier was trained with the binary cross-entropy loss.

RL policy. For our underlying RL algorithm, we implement the Twin Delayed Deep Deterministic policy gradients (TD3) [33]. TD3 was found to be stable, sample efficient and requires few little parameter tuning. Additionally and to further improve sample efficiency, Prioritized Experience Replay (PER) [34] was used. PER alternates between random sampling experience from a buffer and prioritizing experience by replaying important transitions more frequently. For our actor policy, we used the same TCN based architecture of the SL model, to output a 3 unit embedding vector $\tilde{\mathbf{e}}_\mathbf{t}$. Next we concatenate $\tilde{\mathbf{e}}_\mathbf{t}$ and $\mathbf{a}_{\mathbf{t-1}}$ and process it by 2 FC layers with 256 neurons each followed by a Tanh activation. For the critic network, we used a 3 layered FC-NN with 128, 128 and 64 neurons each.

Policy evaluation. Policies are evaluated separately on the train and test objects. Each object is tested for 30 episodes yielding 150 insertion episodes for both the train and novel test objects. If the system fails to insert an object within 15 haptic glances or the object reaches farther than 25 mm\deg from the hole, the episode is declared as failed. Videos of the experiments can be seen in the supplementary material.

B. Contact localization through haptic glances

We begin by analysing the accuracy of contact localization with the three haptic glance primitives described in Section III-B. For each primitive, training data is collected by recording 600 haptic glances per training object yielding a total of 3,000 observations. An additional 600 test samples were collected in a separated session and were not included in the training set. Automated data collection of the robot took an overall time of approximately 13 hours.

Table II presents a comparative analysis of the proposed three haptic glance primitives. The table shows the mean contact localization error, the success rate (over 30 episodes) and average number of haptic glances for each primitive and object, while using (2). The finger perturbations approach

TABLE III
INSERTION PERFORMANCE OF DIFFERENT POLICES

Method		Cyl.	Hex. prism	Ellip. prism	Square prism	Cuboid
Random actions	Success	16%	20%	13%	10%	3%
	#Glances	7.1	5.3	7.9	5.7	6.4
DFC	Success	83%	73%	47%	53%	50%
	#Glances	4.7	5.0	4.5	6.3	5.2
SL	Success	93%	63%	73%	70%	47%
	#Glances	4.5	7.4	3.8	6.9	7.1
RL w/o CL	Success	80%	73%	73%	53%	50%
	#Glances	3.9	4.9	4.8	6.5	5.8
ResRL w/o CL	Success	90%	93%	83%	67%	63%
	#Glances	4.3	4.5	3.7	4.6	4.8
ResRL+CL	Success	97%	93%	90%	93%	90%
	#Glances	3.4	3.1	2.9	3.3	3.5

TABLE IV
INSERTION PERFORMANCE OVER NOVEL TEST OBJECTS

Method		Gum box	Wire roll	Hollow box	Star prism	Flower prism	Inser. time
	∼Clear. ϵ (mm)	1.0	0.25	1.5	2.5	0.5	(sec)
ResRL+CL	Success	93%	77%	83%	87%	80%	40.6
	#Glances	2.0	3.3	3.3	5.2	6.0	
SL	Success	77%	70%	77%	50%	67%	47.6
	#Glances	3.0	4.5	4.2	6.1	5.9	

TABLE V
TEST PERFORMANCE WITH INDUCED $SE(3)$ UNCERTAINTY

Method		Cyl.	Hex. prism	Ellip. prism	Square prism	Cuboid
ResRL+CL	Success	93%	90%	90%	77%	83%
	#Glances	3.1	4.0	3.2	4.1	4.0
SL	Success	87%	80%	73%	53%	70%
	#Glances	5.1	5.9	4.4	6.6	6.0

shows better prediction accuracy both in position and orientation compared to vertical press and arm tilts. Similarly, it exhibits the highest success rate. Observations during the testing have shown that haptic glances that are based on arm movement can result in object jamming. In addition, when slight tilting of the object occurs, it can easily become off-centered with respect to the hole. Therefore, finger motions have benefits for this task as they conform to these tilts. In addition, finger perturbations slightly move the object on the plane with some probability for alignment. Therefore, the approach aid the insertion and increase success rate. It is, therefore, chosen and used in further experiments. The average time for a single finger perturbation haptic glance is approximately 4 seconds.

The insertion classifier was trained with 1,600 recorded haptic glances over the train objects with half `in` and half `out` of the hole. An additional 300 test samples were collected with the test objects. The resulted classification accuracy is 97% and 96% over the train and test objects, respectively. The classifier is further used in the evaluation of RL insertions.

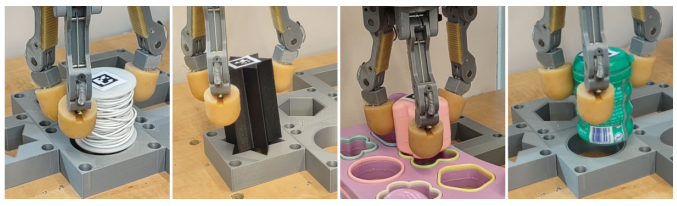


Fig. 6. Snapshots of insertion test trials for the (from left to right) wire roll, star prism, flower prism and gum box.

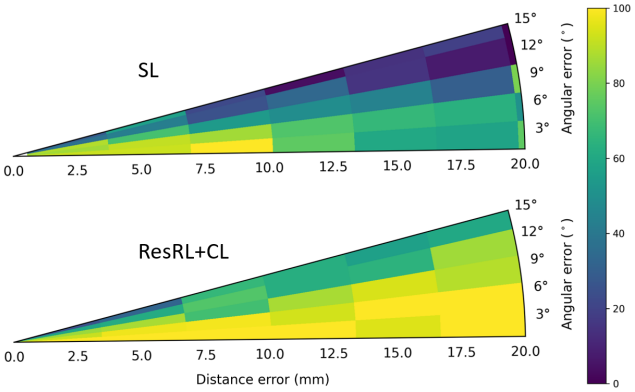


Fig. 7. Heatmap plots indicating the insertion success rate (%) for SL and ResRL+CL with regards to the positional (norm of planar position) and angular (norm of α, γ, θ) object errors encountered during test episodes.

C. RL results

Results above show that SL control is not able to reason about object tilts. Evidently, rectangular objects have the lowest success rate due to tilt sensitivity. Hence, RL-based insertion can acquire unsupervised reasoning for these tilts and act to correct them. RL is, therefore, tested while comparing to baseline methods common in state-of-the-art for rigid hands as discussed in Section II. We compare the ResRL with and without CL to random actions, naive RL, SL and DFC. Direct training of RL policies on real robots may require substantial samples and, therefore, excessive runtime. Hence and following [30], training of the RL policies is accelerated by bootstrapping the actor network with an SL policy trained with 350 samples collected under a random policy. To avoid updates by the untrained critic network, we freeze the actor network for the first 30 episodes of RL training.

Table III presents comparative results between the methods over test episodes on the training objects. First, insertion with random actions lay a baseline for the complexity of the task. ResRL with CL policy acquired the presented results after 320 training episodes (corresponding to approximately 7 hours of robot learning). To ensure a fair comparison, training of each policy is stopped when reached the same amount of episodes. Similarly, SL was trained with the same amount of episodes and, therefore, the difference in results between Tables II and III. Consequently, results reflect the SL requirement for much more training episodes in order to reach performance similar to RL. In addition, DFC alone does not perform well while evidently providing performance boost to ResRL compared to naive RL. The addition of CL to ResRL exhibit the best success rate and lowest number of required haptic glances over all objects. The trained SL and ResRL with CL were further evaluated for insertion of the test objects in Table IV. Since these are everyday objects, hole clearances are varied and reported in the table. Here also, ResRL outperforms SL and exhibits sufficient performance. Although dependent on the manually-set velocity of the arm during pose correction, mean insertion time is reported with clear advantage to ResRL+CL. Note that wire roll and flower prism have a tight clearance and, thus, a lower success rate. Snapshots of some test insertions are shown in Figure 6.

While the above results include $SE(3)$ uncertainties orig-

inating from external contact forces and asymmetric grasps, results yet emphasize ability to overcome them. To further emphasize the ability, we intentionally induce uncertain tilt angles in a test evaluation over the train objects without further policy training. Pitch and roll angles of the arm are uniformly sampled from the range $\alpha, \gamma \sim \mathcal{U}(-10°, 10°)$ during initial object grasp. Table V depicts the evaluation results. In addition, Figure 7 shows heatmap plots of the insertion success rate for SL and ResRL with CL with regards to the positional and angular errors encountered during the test episodes. The results show the ability of ResRL to overcome uncertain tilting angles and large positional errors.

V. Conclusions

In this paper, we have proposed an online-learning framework using residual RL to address the problem of precision insertion with an underactuated compliant hand. With the RL framework, the ability of the hand to correct $SE(3)$ deviations solely using F/T inputs was observed. The learning involves curriculum learning and direct force control to guide and expedite the learning. By exploring the haptic information of a compliant hand during insertion attempts, the proposed method can generalize to various object-hole pairs with high success rate while also overcoming $SE(3)$ uncertainties. While this work discusses haptic-based insertion, the approach can be used in various other applications. For instance, general purpose manipulation in unstructured or occluded environments require robots to accurately understand its scene. Hence, the approach can be fused with vision for increased accuracy. Similarly, in-hand object localization through contact with the environment can be applied. Future extension to the work may reason about the depth of the hole for smooth insertion without drop. Also, additional work may explore transfer learning for hardware generalization.

References

[1] J. Xu, Z. Hou, Z. Liu, and H. Qiao, "Compare contact model-based control and contact model-free learning: A survey of robotic peg-in-hole assembly strategies," *arXiv preprint arXiv:1904.05240*, 2019.

[2] T. Tang, H.-C. Lin, Y. Zhao, Y. Fan, W. Chen, and M. Tomizuka, "Teach industrial robots peg-hole-insertion by human demonstration," in *IEEE Inter. Conf. on Adv. Intel. Mechatronics*, 2016, pp. 488–494.

[3] F. Wang and K. Hauser, "Dense robotic packing of irregular and novel 3d objects," *IEEE Trans. on Rob.*, vol. 38, no. 2, pp. 1160–1173, 2022.

[4] R. R. Ma and A. M. Dollar, "Yale openhand project: Optimizing open-source hand designs for ease of fabrication and adoption," *IEEE Rob. & Aut. Mag.*, vol. 24, pp. 32–40, 2017.

[5] H. Park, J. Park, D.-H. Lee, J.-H. Park, M.-H. Baeg, and J.-H. Bae, "Compliance-based robotic peg-in-hole assembly strategy without force feedback," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 8, pp. 6299–6309, 2017.

[6] A. M. Dollar and R. D. Howe, "The highly adaptive sdm hand: Design and performance evaluation," *The International Journal of Robotics Research*, vol. 29, no. 5, pp. 585–597, 2010.

[7] L. U. Odhner and A. M. Dollar, "Dexterous manipulation with underactuated elastic hands," in *IEEE ICRA*, 2011, pp. 5254–5260.

[8] ——, "Stable, open-loop precision manipulation with underactuated hands," *Int. J. of Rob. Res.*, vol. 34, no. 11, pp. 1347–1360, Sep 2015.

[9] A. Sintov, A. S. Morgan, A. Kimmel, A. M. Dollar, K. E. Bekris, and A. Boularias, "Learning a state transition model of an underactuated adaptive hand," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1287–1294, April 2019.

[10] J. Borras and A. M. Dollar, "A parallel robots framework to study precision grasping and dexterous manipulation," in *IEEE Int. Conf. on Rob. and Aut.* IEEE, May 2013, pp. 1595–1601.

[11] O. Azulay, I. Ben-David, and A. Sintov, "Learning haptic-based object pose estimation for in-hand manipulation with underactuated robotic hands," *arXiv preprint arXiv:2207.02843*, 2022.

[12] B. Wen, C. Mitash, S. Soorian, A. Kimmel, A. Sintov, and K. E. Bekris, "Robust, occlusion-aware pose estimation for objects grasped by adaptive hands," in *IEEE Intl. Conf. on Rob. & Aut.*, 2020.

[13] A. S. Morgan, K. Hang, and A. M. Dollar, "Object-agnostic dexterous manipulation of partially constrained trajectories," *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 5494–5501, 2020.

[14] K. Hang, W. G. Bircher, A. S. Morgan, and A. M. Dollar, "Hand–object configuration estimation using particle filters for dexterous in-hand manipulation," *The International Journal of Robotics Research*, vol. 39, no. 14, pp. 1760–1774, 2020.

[15] A. S. Morgan, B. Wen, J. Liang, A. Boularias, A. M. Dollar, and K. Bekris, "Vision-driven compliant manipulation for reliable, high-precision assembly tasks," in *Robotics: Science and Systems*, 2021.

[16] C. C. Beltran-Hernandez, D. Petit, I. G. Ramirez-Alpizar, and K. Harada, "Variable compliance control for robotic peg-in-hole assembly: A deep-reinforcement-learning approach," *Applied Sciences*, vol. 10, no. 19, p. 6923, 2020.

[17] J. Xu, Z. Hou, W. Wang, B. Xu, K. Zhang, and K. Chen, "Feedback deep deterministic policy gradient with fuzzy reward for robotic multiple peg-in-hole assembly tasks," *IEEE Transactions on Industrial Informatics*, vol. 15, pp. 1658–1667, 2019.

[18] R. Li, R. Platt, W. Yuan, A. ten Pas, N. Roscup, M. A. Srinivasan, and E. Adelson, "Localization and manipulation of small parts using gelsight tactile sensing," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2014, pp. 3988–3993.

[19] S. Kim and A. Rodriguez, "Active extrinsic contact sensing: Application to general peg-in-hole insertion," in *International Conference on Robotics and Automation (ICRA)*, 2022, pp. 10 241–10 247.

[20] W. Haskiya, K. Maycock, and J. A. G. Knight, "A passive compliant wrist for chamferless peg-in-hole assembly operation from vertical and horizontal directions," *Proceedings of the Institution of Mechanical Engineers*, vol. 212, no. 6, pp. 473–478, 1998.

[21] W. Wang, R. N. Loh, and E. Y. Gu, "Passive compliance versus active compliance in robot-based automated assembly systems," *Industrial Robot: An International Journal*, 1998.

[22] D. R. Strip, "Insertions using geometric analysis and hybrid force-position control: method and analysis," in *IEEE International Conference on Robotics and Automation*, 1988, pp. 1744–1751.

[23] Y. Zhang, H. Lu, D. T. Pham, Y. Wang, M. Qu, J. Lim, and S. Su, "Peg–hole disassembly using active compliance," *Royal Society open science*, vol. 6, no. 8, p. 190476, 2019.

[24] J. Ding, C. Wang, and C. Lu, "Transferable force-torque dynamics model for peg-in-hole task," *arXiv preprint arXiv:1912.00260*, 2019.

[25] M. A. Lee, Y. Zhu, K. Srinivasan, P. Shah, S. Savarese, L. Fei-Fei, A. Garg, and J. Bohg, "Making sense of vision and touch: Self-supervised learning of multimodal representations for contact-rich tasks," in *IEEE Inter. Conf. on Rob. and Aut.*, 2019, pp. 8943–8950.

[26] S. Dong and A. Rodriguez, "Tactile-based insertion for dense box-packing," in *IEEE In. Conf. on Int. Rob. & Sys.*, 2019, pp. 7953–7960.

[27] K. Van Wyk, M. Culleton, J. Falco, and K. Kelly, "Comparative peg-in-hole testing of a force-based manipulation controlled robotic hand," *IEEE Transactions on Robotics*, vol. 34, no. 2, pp. 542–549, 2018.

[28] S. Cui, R. Wang, J. Hu, J. Wei, S. Wang, and Z. Lou, "In-hand object localization using a novel high-resolution visuotactile sensor," *IEEE Trans. on Industrial Electronics*, vol. 69, no. 6, pp. 6015–6025, 2022.

[29] R. L. Klatzky and S. J. Lederman, "Identifying objects from a haptic glance," *Perception & Psychophysics*, vol. 57, pp. 1111–1123, 1995.

[30] S. Dong, D. Jha, D. Romeres, S. Kim, D. Nikovski, and A. Rodriguez, "Tactile-rl for insertion: Generalization to objects of unknown geometry," in *IEEE Inter. Conference on Robotics and Automation*, 2021.

[31] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," *arXiv preprint arXiv:1803.01271*, 2018.

[32] L. Pinto, M. Andrychowicz, P. Welinder, W. Zaremba, and P. Abbeel, "Asymmetric actor critic for image-based robot learning," in *Robotics: Science and Systems*, 2018.

[33] S. Fujimoto, H. van Hoof, and D. Meger, "Addressing function approximation error in actor-critic methods," in *International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, vol. 80, 2018, pp. 1587–1596.

[34] T. Schaul, J. Quan, I. Antonoglou, and D. Silver, "Prioritized experience replay," *arXiv preprint arXiv:1511.05952*, 2015.