Reinforcement Learning Grasping with Force Feedback from Modeling of Compliant Fingers

Luke Beddow, Helge Wurdemann, Dimitrios Kanoulas

***Abstract*—Current reinforcement learning approaches for grasping do not consider force feedback combined with compliant grippers, but these features are suited to grocery object grasping, where objects can be bruised by high forces, and exhibit varied weights, textures, and surface irregularities. This work combines force feedback with reinforcement learning on a compliant gripper. We design a three degrees of freedom caging inspired gripper, which can grasp by trapping objects with three compliant fingers and a movable palm. We instrument the fingers with strain gauges for force sensing, then model their bending for simulating grasping at $16.8\times$ real time. Then, we train a reinforcement learning grasping controller based on in-grasp force feedback, with real world transfer achieving 98.0% grasp success rate on training objects, with an average sim2real gap of 3.1%. We demonstrate generalization to 42 novel grocery objects with a success rate of 95.0%, with 80.1% of grasps tolerating a 5N vertical disturbance. In-grasp finger forces averaged 1.4N and palm forces 3.0N. We also validate our method with three finger rigidities, show that our model and in-grasp sensing improve learning and performance, and compare against three baselines.**

***Index Terms*—Grasping & manipulation; Mechanisms, design, modeling & control; AI and machine learning.**

I. Introduction

Robotic grasping of varied of objects, with applications such as logistics or grocery packing, is a significant challenge for which approaches increasingly apply machine learning methods, one of the most prominent being reinforcement learning [1]. Learning methods have in majority been applied to 1 Degree of Freedom (DoF) suction and parallel-jaw grippers [2], [3]. However, such grippers rely on sufficient surface forces for grasping, and most approaches do not include force sensing or feedback, instead determining an appropriate grasping position based on camera data and then executing a hardcoded grasping pattern [4], [5]. These approaches are not suited for applications where contact forces should be limited, monitored, or used as feedback to react during grasp.

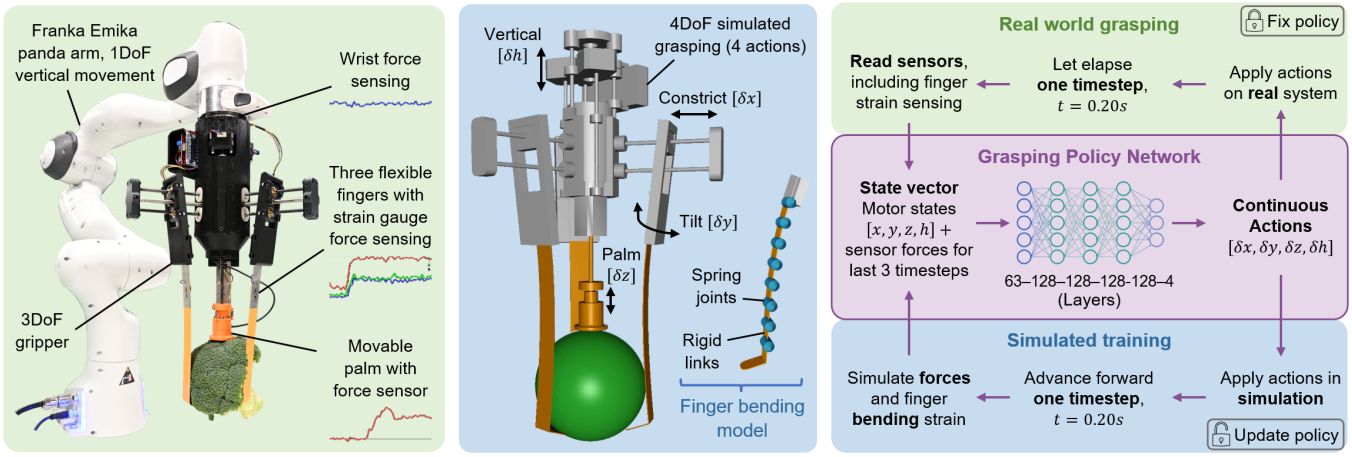
Grocery items, such as fruits and vegetables, have a range of weights, textures, and surface irregularities which can defeat suction [6], [7]. They have a challenging range of geometries, for example grasping both limes and punnets of berries can exceed gripper capabilities [8], and coping with shape irregularities complicates grasping strategies [6], [9]. Groceries can also be bruised and damaged easily, requiring limits on surface forces. Grocery grippers therefore favour compliance or variable stiffness [10] to enable material adaptation to irregular shapes and to inherently limit surface forces.

Reinforcement learning grasping is well suited to grocery grasping, as the structure of incoming observations allows force feedback to monitor and react to surface forces [11], [12]. However, reinforcement learning has not yet been applied to compliant grippers with force feedback, for two key reasons. Firstly, integrating contact sensing into compliant grippers is challenging and remains an active research area [13], as deformable materials make for poor mounting surfaces and the high strain requires equally deformable sensors. Secondly, reinforcement learning approaches usually use simulated training [14], but simulating compliant contacts with accurate forces is computationally complex, making it too slow to use with data-driven simulated training [15]. This complexity is exacerbated by the majority of compliant grippers using nonlinear hyperelastic materials like silicone and relying on friction grasps, which is also challenging to model [15], [16].

The contribution of this work is combining reinforcement learning with force feedback, resulting from a physics-based model of compliant fingers, using caging inspired grasping. We present the significantly updated design of a 3DoF freedom caging inspired gripper from our previous work [17], that uses three fingers and a movable palm which can trap objects inside a cage. We choose a material with linear bending for our three flexible gripper fingers, stainless steel, which allows compliant adaptation to objects as well as contact force sensing via instrumentation with strain gauges. Then, we apply and validate a mathematical model for large deflection finger bending enabling compliant contact physics evaluation in existing robotics simulators at $16.8\times$ real time. We use this to develop a reinforcement learning approach, trained entirely in simulation, that effectively transfers to real world grasping, using force feedback and respecting surface force limits. Our controller learns to favor caging grasps, where objects are surrounded and trapped using geometry, not requiring high surface forces and friction. The result is reliable and stable grasping across a wide range of objects, with our evaluation being done on grocery items. Our approach is outlined in Figure 1. To summarize our contributions:

1) Building on our previous work [17], we significantly redesign a caging inspired gripper with new actuation and added embedded in-grasp force sensing in Section III.
2) A reinforcement learning approach for closed-loop grasping using in-grasp force feedback trained in simulation. The physics-based model of the compliant fingers,

This work is supported by the UCL EPSRC DTP in Fundamental Engineering [2425110] and the UKRI Future Leaders Fellowship [MR/V025333/1] (RoboHike). The authors are with the Departments of Computer Science and Mechanical Engineering, University College London, Gower Street, WC1E 6BT, London, UK.{luke.beddow, h.wurdemann, d.kanoulas}@ucl.ac.uk. Corresponding author: *Luke Beddow*



(a) Real world grasping setup (b) Simulated grasping setup (c) Flowchart of real and simulated grasping using our policy

Fig. 1. Overview of our approach for grasping using compliant fingers and a reinforcement learning policy; a) 3DoF caging inspired gripper mounted on a robot arm, with all five sensors detailed; b) simulated grasping, with arrows indicating the four grasping actions, and demonstration of the finger bending model; c) flowchart showing the difference between real world grasping and simulated training, which both generate grasping actions based on state information.

simulated faster than real-time (Section IV), is combined with the data-driven learning approach (Section V).

3) Validation of: learning and simulation to real (sim2real) transfer across three finger rigidities, dependent on using the bending model and in-grasp sensing; generalization to challenging unseen objects; and comparison against three baseline controllers; all in Section VI.

We review in detail related literature in Section II, and present discussion in Section VII, before a final conclusion.

II. Related Work

This section reviews gripper designs and reinforcement learning approaches, before discussing simulator capabilities and modeling of large deflection cantilevers. For discussion of caging grasping, please refer to our previous work [17].

A. Grippers for Grocery Picking

Commercial logisitics grippers frequently rely on suction grasps, for example Amazon Sparrow, Ocado's Kindred Autograsp, and Righthand Robotics RightPick; the latter two then stabilizing grasps with fingers. However, using suction on fruit and vegetables items is difficult [6], [7], they rarely have flat surfaces and exhibit varied surface texture.

Research into grippers for grocery picking focuses on controlling stiffness, usually leveraging compliant materials which can adapt to irregular shapes and inherently limit contact forces to avoid bruising. A systematic review of 64 current grippers [10] found half used compliance or deformation, including $8/10$ highlighted for delicate or food applications. Compliant elastomer fingers which deform when actuated to curl and surround objects have been presented for delicate handling [28], but required finite element method (FEM) for modeling the complex deformation. The present work proposes flexible steel fingers rather than elastomers, applying a model to characterize their bending in robotics simulators.

Mnyusiwalla et al. [8] benchmarked four grippers for grocery bin picking, all of which used compliance. Grasping success rates in low clutter ranged from $55.6-75.6\%$. Angelini et al. [7] remarked that varied grasping approaches could improve grocery grasping, based on experimental evaluation of their 1DoF gripper, which used the same approach for all objects. The 8DoF DLR CLASH Hand [9], designed for grocery picking, uses antagonistic pairs of motors to achieve variable stiffness. This complexity of actuation enables diverse grasping strategies, but at a cost of control complexity. Hence, reinforcement approaches more commonly use 4DoF grippers or less (see Table I), and higher DoF methods have applied dimensionality reduction [27] or human demonstrations [26].

Integrating tactile sensing into compliant grippers is challenging, as sensors may be required to cope with large deformations, non-rigid mounting, or in-situ manufacture [13]. Highly deformable strain gauges based on liquid metal [29], hydrogel [30], or conductive fibres [31] show promise, but remain difficult to manufacture as well as being prone to hysteresis and nonlinear response. Other sensors for soft robots have similar challenges, such as piezoresitive, capacitive, or pneumatic. Strain gauges are proven and reliable, but are typically used in rigid grippers to isolate relevant forces, hence their application to parallel-jaw grippers, either standard commercial (e.g., Schunk), micro-grippers [32], or 3-axis in grip force sensing [33]. Our approach of instrumenting three flexible steel fingers with strain gauges combines accurate in-grasp force sensing with the benefits of a compliant system.

B. Reinforcement Learning for Grasping

Reinforcement learning allows training grasping controllers which exploit sensor information to actuate grasps, often learning in simulation for time and resource efficiency [14]. Table I compares relevant research to illustrate that no reinforcement learning grasping works combine compliance and closed-loop force feedback. Specifically, we refer to measuring in-grasp forces for continuously adjusting the gripper joints.

Combining tactile sensing and feedback has only been seen on greater than 1DoF grippers, and most frequently without

TABLE I
COMPARISON BETWEEN METHODS USING REINFORCEMENT LEARNING FOR GRASPING.

Author	Year	Gripper	Gripper DoF	Compliant links or joints	Tactile force sensing	Closed-loop gripper feedback	Simulated training
Zeng [18]	2018	RG2 gripper	1	−	−	+	✓
Chen [19]	2022	RG2 gripper	1	−	−	+	✓
Song [20]	2020	RG2 gripper	1	−	−	+	−
Kalashnikov [4]	2018	Parallel jaw gripper	1	✓	−	+	−
Kim [21]	2021	Robotiq 2 finger gripper	1	−	−	+	✓
Liu [22]	2023	4 finger gripper + fingertip suction	2	✓	−	−	✓
Deng [23]	2019	2 finger gripper + moving suction cup	3	−	−	−	✓
Merzic [12]	2019	Barrett hand	4	−	✓	✓	✓
Wu [24]	2020	Barrett hand	4	−	+	✓	✓
Chebotar [11]	2017	Barrett hand	4	−	✓	✓	−
Wu [25]	2019	Barrett hand	4	−	−	−	✓
Kumar [26]	2019	Allegro hand	16	−	+	✓	✓
Liang [27]	2022	Shadow hand	20	−	+	✓	✓
This work	**2024**	**3 finger gripper + moving palm**	**3**	✓	✓	✓	✓

✓ indicates the feature is used, − that it is not. For tactile sensing + indicates using binary contact information, rather than measured forces. For closed-loop feedback + means using feedback with binary open/close grasping, rather than continuous grasping control.

using vision during grasp [11], [12], [24], [26], [27], as in the present work. Merzić et al. [12] used contact forces with trust region policy optimization, but trained and tested on five objects in simulation only without considering force limits. Chebotar et al. [11] used relative entropy policy search, using BioTac sensors for force feedback. However, they learned a separate policy for each of their three training objects, then a master policy to select between them and handle regrasping. Our work approaches the problem differently with an end-to-end view of training one network that should generalize to different objects as well as handle regrasping implicitly.

Wu et al. [24] presented multi-fingered adaptive tactile grasping (MAT), using modified proximal policy optimization (PPO) with the Barrett Hand, and achieving 98.7% real grasp success rate. In contrast to our approach, they used binary sensing without force limits or compliance during feedback grasping; however, the high success rate and generality of their multi-fingered approach make MAT the most suitable related work to apply on our gripper. We therefore used MAT as a baseline to compare against our method, and Section V-C details further the technical aspects.

Liu et al. [22] presented a similar compliant modeling approach. They used a 2DoF soft gripper trained in simulation with double deep Q-learning, based on pre-grasp depth images. They modeled their variable stiffness tendon actuated gripper fingers with circular arcs then achieved bending in simulation by splitting each finger into four segments and directly setting the joint angles between them, based on a target pose. Then, they assumed the fingers will behave rigidly. In contrast, our work models finger compliance during interaction. This is essential for our force feedback grasping, as contact forces need to be accurately modeled to enable real world transfer.

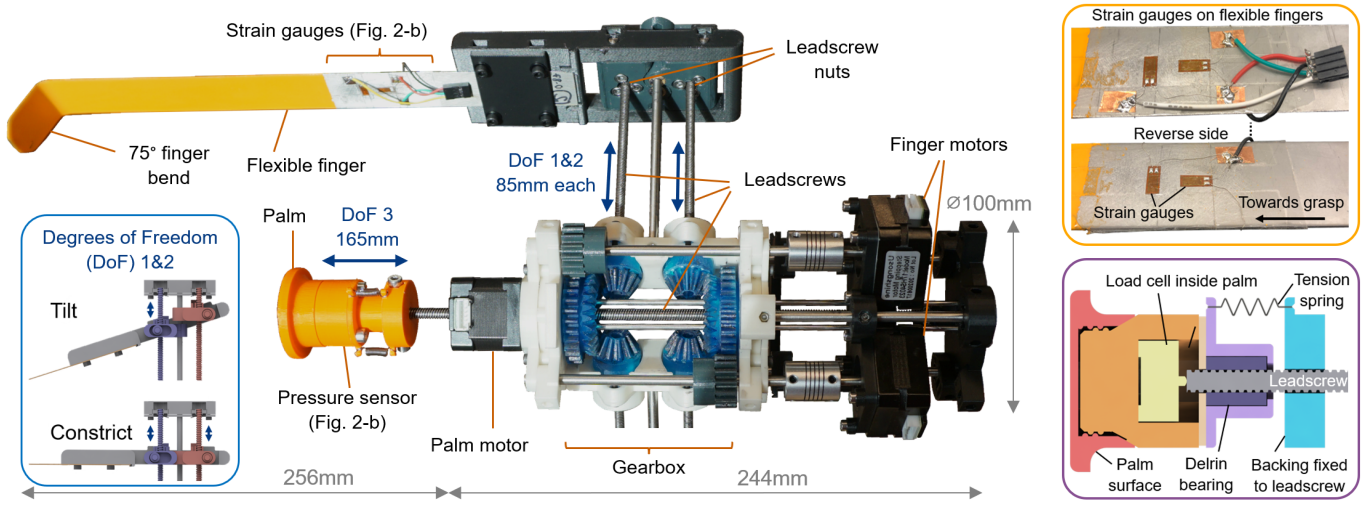
C. Simulating Grasping and Deformation

The gold standard for deformation simulation is commercial FEM software such as Abaqus, ANSYS, and COMSOL, which are too slow for reinforcement learning [15]. Instead, simulators for robotics use simplified models, such as mass-spring models with volume constraints (MuJoCo, CoppeliaSim, PyBullet), neo-hooken volumetric FEM (PyBullet), or position based dynamics (Nvidia Issac). Nvidia FleX offers co-rotational linear volumetric FEM for higher fidelity analysis of stress and strain fields, however timesteps must remain small (e.g., 0.67ms) and framerates low (e.g. $5-10$ *fps*) [34]. This results in simulation $10-100\times$ slower than real time. Simulation of deformation is integrated with traditional rigid-body physics for robotics in FleX, making it suitable for grasping [26], [34]. The computational expense remains significant for reinforcement learning; via a tailored model the present work simulated bending at $16.8\times$ real time in MuJoCo.

Accurate physics simulation of contact mechanics and tribological effects such as friction is an ongoing area of research [15], [16]; especially when considering nonlinear hyperelastic compliant materials like elastomers. The current work limited surface forces, avoided elastomers, and prioritized geometry to secure objects with a caging approach, which all reduce reliance on accurate simulation of friction.

D. Modeling Large Deflection Cantilevers

Analytical solutions for large deflection cantilevers have been presented, but vertical point end load requires numerical methods to solve [35]. Simplified pseudo-rigid-body models, based on dividing a beam into N spring jointed segments, have been presented to match tip deflection [36], where numerical optimization determines parameters; typically, link lengths and spring stiffnesses. Su [37] presented an $N=3$ method with linear spring stiffness, Vedant and Allison [38] presented a method suitable for N joints, using nonlinear spring stiffness. Pseudo-rigid-body models have been applied to robots, such as an $N=6$ model for a continuum manipulator [39] and an $N=3$ catheter model [40]. Roesthuis and Misra [41] presented a rigid link model, deriving a joint stiffness to connect N segments for a continuum manipulator. They applied the model in a control loop for an actuated prototype, based on known applied forces. We also use a rigid link model, but apply it for computationally efficient data-driven training of grasping, adapting it for integration in a physics simulator.



(a) Gripper section view

(b) Finger and palm sensors

Fig. 2. Design of the gripper: a) Section view showing the mechanical design, sensor placements, and dimensions, with blue arrows indicating degrees of freedom; b) Finger sensing with strain gauges mounted in a full-bridge (top), palm cross-section showing load cell which measures axial forces (bottom).

III. Gripper Design

We summarize the grasping principle, and then present the 3DoF gripper, which has entirely new actuation, and embedded finger and palm sensing, compared to our previous design [17].

A. Caging Inspired Grasping Principle

The gripper design results in two mechanisms in the grasp. Firstly, the caging effect, surrounding and trapping objects between the fingers and palm. Secondly, friction, mediated by finger stiffness as well as opposition between the fingers and the palm. Caging works well for spheres, which approximate many fruits and vegetables, and finger flexibility aids adaptation to irregular objects. Friction is required for objects with parallel edges, such as sideways cylinders, which cannot be fully caged by the design; it also helps secure grasps, as the flexible fingers can be loaded like springs to a desired force.

B. Actuation System and Mechanical Design

The gripper uses three stepper motors, with all motions driven via non-backdrivable leadscrews, shown in Figure 2-a, to prevent the cage opening from disturbances. The three fingers perform identical 2DoF motions, driven through a gearbox. Each finger is connected to two leadscrew nuts via a pin joint and a sliding pin joint. Synchronized leadscrew motion causes constriction, the finger displaces whilst maintaining a constant angle to surround objects. Whereas, differing leadscrew actuation will tilt the finger through different angles, to get underneath objects and to more directly mediate finger bending. The maximum achievable tilt angle is $\pm 40°$, allowing the fingertips to touch via tilt only from the maximally open position. Without tilting, the fingers can open to a maximum grasp diameter of 275mm, and a minimum of 105mm. Hence, the largest objects graspable fit within the 275mm grasp diameter, and the smallest is approximately a 40mm diameter sphere (e.g. strawberries). The palm, which extends and retracts up to 165mm to cage objects and press them into the fingers, is driven by a non-captive stepper motor where the leadscrew shaft passes through the motor. The finger motors both have torque of 130Nmm, reliably exerting grasp forces up to 3N, whilst the palm motor has a torque of 100Nmm, able to exert up to 30N. The gripper weighs 2.2kg. The motors, driven up to 400rpm, can cover their workspace within 5s.

Compared to the previous gripper [17], the new actuation has reduced torque requirements, less backlash and wear, decoupled tilt and constriction, and the ability to open the fingers past parallel. Weight reduced by $2\times$, volume by $5\times$.

The gripper fingers should adapt to objects whilst being feasible to model efficiently. 304 stainless steel was selected, fulfilling the criteria: 1) linear elastic bending under grasp forces; 2) food safe; 3) suitable for instrumentation with strain sensing; 4) Youngs Modulus resulting in appropriate deflection under grasp forces (see Section IV-B). Lower yield strength compared to other materials was a limitation, as well as possibility of long term fatigue failure. A 75° bend, 35mm long, was placed at the end of each finger, and no high friction material like silicone was added, as caging rather than friction was intended to be the primary mechanism for grasping.

C. Embedded Sensors and Electrical Design

The gripper used force sensing integrated into three areas: the fingers, palm, and wrist. The gripper was mounted onto a Franka Emika Panda robotic arm, which provided the external vertical forces at the wrist. Measuring these out of grasp forces intended to allow detecting of collisions and grasped object weights. Finger and palm sensing of in-grasp forces was achieved onboard, as shown in Figure 2-b, aiming to sense object geometries and allow grasping using force feedback.

Finger force sensing used four strain gauges wired in a full bridge configuration. Each side of the finger had two gauges, one mounted axially to measure bending and the other mounted perpendicularly to cancel out Poisson effects, as shown in Figure 2-b. Strain was linear from an end applied

force (see Section IV), and large enough that gauge placement was not critical. All fingers used during experiments (nine) were calibrated with loads up to 3N. The largest finger sensor uncertainty was ± 38mN, at the 95% confidence interval (CI).

The palm sensor measured axial compressive force using a penny load cell, with 30N maximum allowable force, as shown in Figure 2-b. A delrin plastic bearing allowed axial force transmission and hence measurement, but protected from out of axis forces. Contact between the load cell and leadscrew was preloaded by three equally spaced tension springs to eliminate backlash and overcome stiction. The palm was calibrated up to 5N and had uncertainty ± 39mN (95% CI).

An onboard Arduino GIGA was used to control the motors, read the finger and palm sensors at 20Hz, and send and receive data at 20Hz via wired USB-C. Power input was 24V 5A.

IV. Modeling Beam Bending with Segments

This section applies a rigid link approximation to model finger bending in MuJoCo, then validates the accuracy with simulated and real world experiments.

A. Model for a Segmented Beam

The flexible fingers were modeled using a rigid link approach [41], being divided into N equal segments, connected by revolute joints with linear stiffness. Joint stiffness depended on the beam rigidity, which is the product EI of the Young's modulus, E, and the cross-sectional 2^{nd} moment of area, I. The flexible fingers were considered as cantilever beams under a point end load. By assuming negligible mass for a thin beam and small angles, virtual work gives a static equilibrium expression for segment angle relative to the previous segment, θ_n, in a beam divided into N segments connected by spring joints with stiffnesses k_n:

$$\theta_n = \frac{N-n+1}{N} \cdot \frac{FL}{k_n} \tag{1}$$

where n is the segment number (from 1 to N segments), F is the applied end load, and L is the total unbent beam length.

We assumed an Euler-Bernoulli beam, and that the small deflection approximation was valid despite the fact we wanted to model large deflections. Bisshopp [42] showed that this assumption is appropriate and introduces very little error in large deflection cases provided that $FL^2/EI < 1$. This was valid in our case (max 0.95 at 5N). The beam deflection, $\delta(s)$, at the n^{th} joint (assuming positions $s_n = (n-1)L/N$ along the beam) is given by small deflection theory. Then, the relative angles, θ_n, of segments that connect these deflected positions can be determined using the gradient:

$$\theta_n = \frac{\delta(s_{n+1}) - \delta(s_n)}{s_{n+1} - s_n} - \frac{\delta(s_n) - \delta(s_{n-1})}{s_n - s_{n-1}} \tag{2}$$

$$\theta_n = \begin{cases} \dfrac{FL^2}{2EIN^2}(N - \frac{1}{3}), & n = 1 \\ \dfrac{FL^2}{EIN^2}(N - n + 1), & n > 1. \end{cases} \tag{3}$$

We can rearrange Equation (1) and substitute in our expression for θ_n from Equation (2) to get our relation of beam rigidity, EI, to joint stiffness, k, at the n^{th} joint:

$$k_n = \begin{cases} \dfrac{2EIN}{L}\dfrac{N}{N-\frac{1}{3}}, & n = 1 \\ \dfrac{EIN}{L}, & n > 1. \end{cases} \tag{4}$$

The final expression describes the choice of joint stiffness k_n which should be set in simulation for modeling finger bending. This expression differs from Roesthius and Misra [41], but will converge to their result with sufficiently high N. The additional $N/(N-\frac{1}{3})$ term, arising from our derivation based on point end load, can be considered a point end load correction factor for low N.

B. Model Validation Using Theory and Real Experiment

The joint stiffness model was evaluated using segmented beams in MuJoCo, then validated against theory and real data. Segmented beams are convenient to create in Unified Robotics Description Format (URDF) format, being a series of identical links connected by revolute joints, with the torsional stiffness given by Equation (4). Beams with 3 to 30 segments were generated in URDF for this validation, with three different beam rigidities, $EI = [0.29, 0.34, 0.40]$ Nm^2, corresponding to the gripper fingers tested in grasping experiments (Section VI)[1].

Validation was done for a point end load with real experimental data; and for a point end load, uniformly distributed load, and point end moment compared against theory. For validation with a point end load, three end masses were applied: 100g, 200g, and 300g; chosen to keep stress below yield. For validation with the uniformly distributed load and point end moment, three loads were again applied, scaled to result in equal deflection to the three mass conditions for point load. The deflected shape of the real beams was extracted by hand from camera images. Hanging masses were used to apply end loads to horizontally clamped beams, then additional deflection due to beam weight was removed by subtracting the unloaded deflection shape from each loaded deflection shape.

1) Error metric: The percentage area error was used to quantify model error, referring to the area of the errors between two curves divided by the total area under the reference curve (theory or real data). The tip position error was verified to be in agreement, being on average 0.34% less than area error.

2) Theory validation results: Figures 3-a, 3-b, and 3-c show that errors between the model and theory are low. For $N \geq 8$, the maximum error was 2.9%. The model converges as N increases, but does not converge to zero error, resulting in some small error increases with N. This shows our assumptions introduced up to 3% error compared to theory.

3) Real experimental data validation results: Figure 3-d shows that real experimental data errors are comparable with the theory case (Figure 3-a), however maximum error increased to 5.3% rather than 2.9%. Despite increased variance

[1] $I = t^3 w/12$ for finger thickness, t, and width, w. We vary EI using real finger geometry $t \times w$ (all mm): 0.86×28.0; 0.96×24.0; 0.96×28.0.

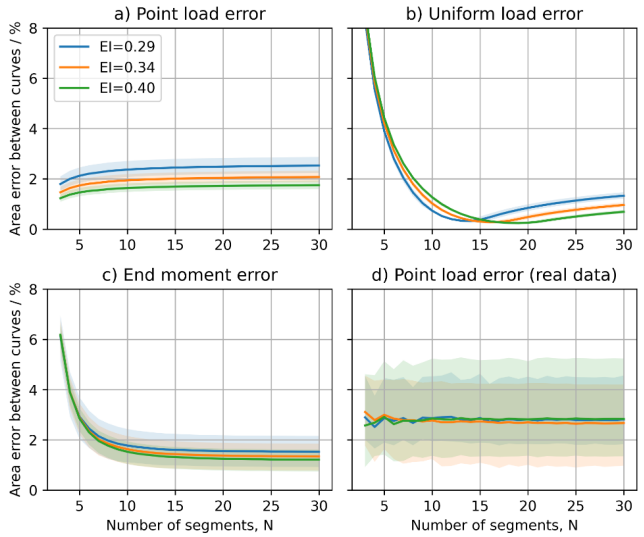


Fig. 3. Model error with respect to theory (a, b, c) and real data (d). The error is averaged across three applied loads of 100/200/300g (or equal maximum deflection in b) and c)), with the range shown by the shaded region.

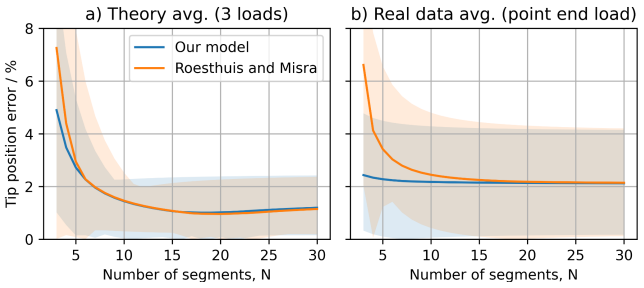


Fig. 4. Averaged (avg.) errors over the three load magnitudes for each of the three rigidities, comparing our model against Roesthuis and Misra [41]: a) Averaging the errors with respect to the three theory load cases, point end load, uniformly distributed load, and end moment; b) Error with respect to real data, for point end load. Range shown by the shaded region.

(min/max values shown by the shaded region), the convergence was similar, with an average difference between all individual datapoints of only 1.5%. This additional uncertainty may be from imperfect camera calibration, measurement error during hand extraction of bending profiles, and selection of Youngs Modulus ($E = 193$GPa [43] was used for stainless steel).

4) Comparison against Roesthuis and Misra: Figure 4 compares our model against the rigid link model from Roesthuis and Misra [41], using tip position error to align with their validation approach. The comparison against theory considers all three load cases, where error varied by only 0.1% on average, but maximum error (shown by shaded region) for our model averaged 2.7% less for $N < 10$. Our model had faster convergence for point end load, and theirs for end moment. Our model had lower point load error across all N using real data, averaging 1.3% less for $N < 10$.

5) Discussion: The validation illustrates that the model is in agreement with both theory and real measurements, with error below 5.3% for real data and 2.9% for theory, for $N \geq 8$. The results in this section demonstrate that the model converges as the number of segments used to approximate the beam increases, but does not converge to zero error. The load conditions, beam rigidities, and force ranges included in this validation cover our requirements for simulating grasping. Point load is the most frequently expected case, but the good performance on uniform loads and end moments provides confidence the model is suited to more complex cases. We showed our model converged faster for point end load on the actual gripper fingers, and had reduced maximum errors for small N compared to Roesthuis and Misra. Since computational efficiency favors small N, this justified using our model during simulations.

We selected $N = 8$ for simulating grasping as larger N yields only marginal accuracy improvements whilst also slowing down the simulation, for example $N = 9$ is 20% slower. More joints have greater computational expense and smaller segments with tiny inertia values cause simulator instability, necessitating smaller physics timesteps as segments are insufficiently "damped" by lack of inertia. We achieve a speed up factor 2.2 by scaling segment inertia by 50, with no difference to model static equilibrium and no measurable difference to grasping, thus running simulation at $16.8\times$ faster than real time. The mujoco timestep used was 2.6ms. Overall, model error was considered acceptable given simulation already introduces a sim2real gap, and supported by the simulated and real grasping results (Section VI) empirically demonstrating effective sim2real transfer.

V. Learning and Evaluating Grasping

Proximal policy optimization [44] was used to train a grasping controller in simulation, which was then evaluated with two experiments in the real world.

A. Reinforcement Learning Formulation

We formalize grasping as a Markov Decision Process (MDP) defined by $\langle \mathcal{S}, \mathcal{A}, p, \gamma \rangle$, with observation space $\mathcal{S}$, action space $\mathcal{A}$, a joint probability $p(r, s'|s, a)$ of a reward r and next state s', given state s and action a, and discount factor $\gamma \in [0, 1)$. A policy π selects a to maximize the expected return, $\mathbb{R}$, according to the reward function $r : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$. Any policy π has associated an on-policy value function $V^{\pi}(s)$, giving $\mathbb{R}$ from following π given starting state s. The advantage function $A^{\pi}(s, a)$ describes the relative change in $\mathbb{R}$ from a particular action a compared to a random action.

PPO is a model-free policy gradient method, where the gradient of policy performance with respect to $\mathbb{R}$ is maximized. A fixed policy π_{θ_k}, parameterized by a neural network, collects trajectories $\mathcal{D}_k$ over T timesteps. Then, we use the PPO-clip objective, maximizing with stochastic gradient ascent:

$$\theta_{k+1} = \frac{1}{|\mathcal{D}_k|T} \sum_{\tau \in \mathcal{D}} \sum_{t=0}^{T} \min \Big(P_\theta A^{\pi_{\theta_k}}(s,a),\ g\big(\epsilon, A^{\pi_{\theta_k}}(s,a)\big) \Big) \tag{5}$$

$$\text{where:} \qquad P_\theta = \frac{\pi_\theta(a|s)}{\pi_{\theta_k}(a|s)} \tag{6}$$

$$g(\epsilon, A) = \begin{cases} (1+\epsilon)A & A \geq 0 \\ (1-\epsilon)A & A < 0 \end{cases} \tag{7}$$

The hyperparameter ϵ clips optimization step size, and was set to 0.2. The advantage $A^{\pi_{\theta_k}}(s,a)$ was calculated with generalized advantage estimation [45], based on $\hat{R}_t$, the rewards from $t \to T$, and a value function network, parameterized by ϕ, which was updated using gradient descent:

$$\phi_{k+1} = \arg\min_{\phi} \frac{1}{|\mathcal{D}_k|T} \sum_{\tau \in \mathcal{D}_k} \sum_{t=0}^{T} \left(V_\phi(s) - \hat{R}_t \right)^2 \quad (8)$$

B. Applying Learning to Grasping

The grasping proceeds with observations of the current state $s_t = \{s_t^{motors}, s_t^{tactile}\}$ containing motor state information s_t^{motors} and sensor information $s_t^{tactile}$, which is a 63 element vector. All values are normalized between $[-1, +1]$.

s_t^{motors} concatenates vectors for the 4DoF, these being the three gripper motors denoted $[x, y, z]$ and the gripper vertical height $[h]$, as shown in Figure 1. Each of these vectors has 7 elements and is structured as $\{a_{t-3}, b_3, a_{t-2}, ..., b_1, a_t\}$ where a_{t-i} is the position at the specified timestep, up to $i = 3$ into the past, and where $b_i = sign(a_{t-i+1} - a_{t-i})$, i.e., the sign of the change in the position, taking only values $b_i = [-1, 0, +1]$.

$s_t^{tactile}$ concatenates vectors for the five sensors: three finger sensors, the palm, and the wrist. Each of these vectors has 7 elements and is structured as $\{c_{t-3}, d_3, c_{t-2}, ..., d_1, c_t\}$ where c_{t-j} is the sensor reading at the specified timestep, up to $j = 3$ into the past, and where d_j is the average of all sensor readings received in-between timesteps, typically three readings.

The action at the current timestep, a_t, is a 4 element vector corresponding to the change in motor state of each DoF. Elements are clipped to remain between $[-1, +1]$ then multiplied by step sizes $[x{*}2\text{mm}, y{*}0.015\text{rad}, z{*}4\text{mm}, h{*}2\text{mm}]$. We used MuJoCo for simulating the gripper, approximating the parallel leadscrew actuation with a revolute and prismatic joint.

A learning rate of 5×10^{-5} was used with the Adam optimizer [46] ($\beta_1 = 0.9, \beta_2 = 0.999$). Each training epoch had 6000 samples, with updates iterated 80 times. γ was 0.99, λ was 0.97, target KL divergence was 0.01, and maximum KL ratio 1.5. Uniformly random action noise ± 0.05 was added.

The key reward signals were $+1$ and episode complete for a successful grasp, -1 and episode terminated for an object out of bounds, and a -0.004 penalty for each step, the reciprocal of the step limit, $T = 250$. A successful grasp required lifting the object whilst all contact forces remained in desired ranges (exact criteria given in upcoming Section V-C). A sparse reward failed to learn, particularly as achieving the desired palm force required many steps before contact. Therefore, a shaped reward based primarily on sensor forces was used, both to guide learning and to teach force limits. Each of the three sensors (fingers, palm, wrist) had a desired force range, $r_1 \leq f < r_2$, which incurred a reward of 0.0008, a dangerous force range, $r_2 \leq f < r_3$, which incurred a penalty of -0.0008, and a termination threshold, $f \geq r_3$, above which incurred a penalty of -1 and ended the episode. For each finger and the palm sensor $(r_1, r_2, r_3) = (1, 4, 5)$N and for the wrist $(r_1, r_2, r_3) = (6, 6, 8)$N. Rewards of $[0.0008, 0.0016]$ were given for lifting the object $[> 0, > 15]$mm off the ground.

1500 simulated objects were used for training, with 100 reserved for testing. The set was composed of 13 elementary object categories: spheres, sideways cylinders, upright cylinders, various cuboids etc. Within each category, numerous objects were generated varying in size between the smallest and largest the gripper should grasp. For example, sphere diameter varied from $50 - 160$mm. Objects also had varying: edge fillet radius, from $5 - 45$mm; friction coefficient, from $0.5 - 2.0$; and density, from $100 - 300$kg m^{-3}. The average mass was 145g, maximum 500g. An illustration of the training set can be seen in Figure 5-a, which is a real life recreation of the simulation test set, reduced from 100 to 30 objects.

The central idea in using elementary object shapes was to generalize to irregular shapes, for example a sphere generalizing to an apple, by using noise during training. Each episode, every incoming sensor reading was transformed by a uniformly random static offset $u_1 \in [-0.05, 0.05]$, and a random Gaussian noise with standard deviation, $s = 0.025$; every motor state reading was transformed by a uniformly random static offset $u_2 \in [-0.025, 0.025]$. These were applied following normalization of all readings between $[-1, +1]$, with saturation enforced. This means two identical objects would be associated with noise with respect to both sensor readings and gripper motor states. This noise, when combined with the flexible fingers adapting to objects shapes, aimed to allow generalization from elementary shapes to irregular ones.

C. Training and Experimental Protocol

During simulated training, simulated experiments, and real world experiments, the following definitions were consistent:

1) Grasping Trial: Objects were placed under the gripper aligned with the center of grasp, with ± 15mm position noise, in any random rotation which did not collide with the initial position of the fingers. The gripper was initialized to a height $h = 10$mm above the table, with ± 5mm of noise. Trials ended after 250 steps, if the object went out of bounds, or following automatic detection of a successful grasp (verified in real life).

2) Successful Grasp Criteria: a) object picked up off the ground; b) average finger force exceeds 1N; c) palm force exceeds 1N; d) no individual finger or palm force exceeds 4N; e) height $h > 20$mm. The success rate percentage (SR %) was given out of the total number of trials. The 1N minimum ensured stable caging grasps. The 4N limit prevented high forces, and compromised between real fingers actuating most reliably up to 3N, and 1N overhead which improved learning.

3) Stable 5N Grasp Criteria: a) successful grasp established; b) object resists 5N vertical disturbance and remains grasped. The palm would push the object directly downwards and out of grasp (up to 30mm) whilst measuring whether the force required exceeded 5N, as shown in Figure 6-a. This metric, "Stable 5N %", was a percentage out of the number of successful grasps, and was only evaluated in real life.

Figure 5 shows the three object sets used for real world evaluation. Experiment One compared simulation success rates with real life, recreating the simulation test set from a representative set of polystyrene objects painted with a hard coat. The real masses were less (average 35g) than the simulation masses (average 145g). Experiment One had four objectives: compare the sim2real gap; compare performance on three

(a) Real world version of simulation test set (30 objects)

(b) YCB grocery items [47] (18 objects)

(c) Real grocery items (24 objects)

Fig. 5. Object sets used for experiments, with mass distributions indicating object masses throughout each set. Objects shown in grasping orientation. For scale: spheres in a) range from $50-160$mm diameter; the strawberry in b) is a 45x55mm ellipsoid, the red cracker box is 210x160x60mm; the lime in c) is a 50x60mm ellipsoid, the salad bag is 210x210x55mm.

finger rigidities ($EI = 0.29, 0.34, 0.40$); compare with and without using the mathematical model (Section IV) at training time; and an ablation study from zero sensors to all three sensors (fingers, palm, wrist) to assess the impact of each sensor. 5 trials were performed on each of the 30 objects.

Experiment Two evaluated generalization to two sets of unseen grocery objects, the 18 YCB food objects [47], and a set of 24 real grocery objects, with 5 trials per object. Real groceries were chosen to range across a variety of weights ($93-638$g), sizes (50mm diameter lime to 180mm long cartons), geometries (spheres, ellipsoids, cylinders, cuboids, irregular shapes), and dynamics (single items, bunches, bags).

We compared our PPO policy against three baselines. Firstly, a heuristic strategy: 1) detect the table using the wrist sensor; 2) tilt fingers to a $15°$ angle, so fingertips are parallel to the table; 2) constrict the fingers until the finger forces averaged 1.5N; 3) lower the palm until the force exceeded 1.5N; 4) lift up, using feedback control to maintain the finger and palm forces within the limits for a successful grasp. These force values were chosen to result in similar forces to the PPO policy at the end of grasp (see Section VI-B), and to ensure grasps on the test objects (grasped $40/42$ during experiments).

A second baseline controller was trained using MAT [24]. Originally presented for the Barrett hand, the state vector

TABLE II

EXPERIMENT ONE: GRASPING SUCCESS RATE (SR %) IN SIMULATION (SIM.) AND ON REAL VERSION OF SIMULATION TEST SET (FIGURE 5-A).

Use model	Sensors in use	Finger rigidity (Nm^2)	Sim. SR %	Real SR %	Real 5N stable %
Yes	All	0.29	91.9	93.3	76.4
Yes	All	0.34	91.5	98.0	93.2
Yes	All	0.40	89.2	90.7	92.7
No	All	0.29	81.5	61.3	46.7
No	All	0.34	81.5	63.3	45.3
No	All	0.40	81.5	66.7	59.0
Yes	Fingers, palm	0.29	88.4	79.3	51.3
Yes	Fingers	0.29	64.3	80.0	58.4
Yes	None	0.29	61.6	68.7	35.9

TABLE III

EXPERIMENT TWO: GRASPING SUCCESS RATE (SR %) AND SUCCESSFUL GRASP FORCES ON GROCERY OBJECTS (FIGURE 5-B,C).

Controller	Grocery objects	Avg. finger force (N)	Avg. palm force (N)	SR %	Stable 5N %
PPO (ours)	YCB	1.29	3.08	**94.4**	**80.1**
Heuristic	YCB	1.49	2.74	83.3	53.3
MAT + ours	YCB	1.73	2.69	76.7	44.9
MAT [24]	YCB	1.36	0.00	84.4*	40.5
PPO (ours)	Real	1.42	2.98	**95.8**	**80.1**
Heuristic	Real	1.49	2.57	78.3	47.3
MAT + ours	Real	1.72	3.21	80.1	61.9
MAT [24]	Real	1.27	0.00	75.0*	53.3

*Successful grasp only required lifting the object, as in [24].

reduced from $15,288$ values to 779, as our gripper has 3 joint angles compared to 8, and 5 force sensors rather than 96 tactile cells. We used continuous force readings (not binary), to encode more information, and trained using the same elementary objects and state reading noise as our main method. The fingertip height was fixed at 5mm above the table, as MAT has no vertical height action ($\pm h$). Learning did not occur using the force-dependent successful grasp criteria, so for this baseline a successful grasp was defined as lifting the object.

A combination of MAT and our method was used in a third baseline, which did use the full successful grasp criteria. We changed to our reward function, our 4DoF actions (with $\pm h$), and no curriculum. We retained MAT's state vector, network architecture, Bernoulli action sampling, and loss function.

VI. RESULTS

Results are given for Experiment One, comparing simulated vs real performance; and Experiment Two, which evaluated generalization to grasps of YCB and real grocery objects.

A. Experiment One: Training Objects

The results for Experiment One are shown in Table II. The top three rows compare policies trained using the mathematical model for finger bending (Section IV), applied to three different finger rigidities. Simulation success rate peaked at 91.9% and varied by only 2.7%, demonstrating effective learning between rigidities using the model. The best real world success rate was 98.0% with $EI = 0.34$, with 93.2% of these grasps

(a) Grasp criteria

(b) Successful grasps of grocery items

Fig. 6. Grasps achieved using our PPO policy; a) Illustrating the two criteria used to evaluate grasps: a successful grasp requires a lifted object and safe finger and palm forces ($1-4$N), whereas a stable 5N grasp also requires tolerance to a 5N vertical disturbance from the palm, which pushes downwards up to 30mm whilst measuring whether the required force exceeds 5N; b) Grocery item grasps, real in the top two rows and YCB [47] in the bottom row.

resisting a 5N disturbance. All three policies transferred well, with an average sim2real success rate improvement of 3.1%.

The middle three rows in Table II illustrate training with rigid gripper fingers, and not using the finger bending model. Simulated success rates only reached 81.5%, a drop of 10.4% compared to training with finger bending. Sim2real transfer was worse using this policy, with average performance reduction of 17.7%, resulting in a best real success rate of 66.7%.

The bottom three rows in Table II show an ablation reducing the number of sensors available at training and testing time from all (fingers, palm, wrist) to none, with $EI = 0.29$. As each sensor was removed, the average performance reduction was 10.1% for simulated success rate, 8.2% for real success rate, and 13.5% for 5N stable grasps. Policy transfer was more variable than with all sensors, with a 10.3% sim2real gap.

B. Experiment Two: Generalizing to Grocery Objects

The results for Experiment Two are shown in Table III, comparing generalization of the best PPO policy from Experiment One (98.0%, all sensors, $EI = 0.34$) to unseen grocery objects, against three baseline controllers (detailed Section V-C). Our PPO controller achieved 94.4% success rate on the YCB grocery set, and 95.8% on the real grocery objects. The stable 5N grasp rate was 80.1% on both object sets. Our PPO controller performed 15.0% better than the best baseline for successful grasps (80.0%, heuristic), and 25.3% better than the best baseline for stable 5N rate (54.8%, MAT+ours). The three baselines varied in success rate by only 2.1% and stable 5N rate by 7.3%. Our PPO policy averaged the same total in-grasp force, 7.1N (three fingers plus palm) as the heuristic, whereas MAT+ours averaged 8.2N, and MAT averaged 3.9N.

VII. Discussion

Experiment One demonstrated very successful learning and sim2real transfer, reliant on our model for compliant finger bending and in-grasp sensing. Experiment Two showed this learning generalized to novel and challenging grocery objects with high grasp success rates and grasp stability, effectively combining with force feedback and compliant gripper design.

A. Experiment One: Training Objects

The 98.0% highest grasping success rate in Experiment One demonstrated excellent grasping reliability, whilst 93.2% of these grasps then resisting a 5N vertical disturbance showed strong grasp stability. The policy used the wrist sensor to keep the fingertips close to the table and slip them underneath objects. Once in grasp, motions became smaller to keep forces in range and prevent smaller objects falling out of grasp before the fingers closed. The policy learned to tilt the fingers only through small angles, keeping the fingertips from going past parallel to the ground ($\leq 15°$) and reducing bending; consequently, limiting surface forces on the object.

The finger bending model and training method worked across all three finger rigidities. The average sim2real difference was 3.1%, always improving in the real world. The main sim2real difference was in palm contact forces, which in real life increased more sharply following contacts, compared to smoother force signals from MuJoCo's soft constraint physics. Hence, real policies exhibited more reactive palm behavior, frequently retreating after initial contacts with more rigid objects. This sim2real difference in contacts was not seen with the fingers as their compliance inherently smoothed forces.

The most common cause of failed trials in both simulation and real grasping was objects which were grasped and picked up, but did not qualify as successful grasps because the forces were outside the acceptable $1-4$N limits, usually finger forces too low or palm forces too high. The $EI = 0.34$ policy best solved this problem in real life with feedback adjustments and improved fine motor control, leading to the best success rate.

Training without the bending model, instead using rigid fingers, resulted on average in: 9.4% lower simulation success

rate; 30.3% lower real grasping success rates; and 20.8% worse sim2real transfer. Real success rates and sim2real transfer reduced with less rigid fingers, which had a wider gap from the rigid fingers during training. These results demonstrate that training with the bending model was essential to our approach.

The ablation study demonstrated that in-grasp sensing with our embedded sensors improved performance. Using fewer sensors, policy sim2real transfer was more variable. The fingers and palm policy overfit and transferred poorly to the real world, due to observed over-reliance on palm force signals which had the biggest sim2real difference. The fingers only and no sensors policies transferred well due to a slow, cautious approach. They closed the grasp very gradually, relying on the automatic detection of successful grasps (which did use the sensors). Training was harder with fewer sensors, the number of training runs able to "learn" and achieve SR $> 5\%$ reduced from $14/15$ with all sensors to $1/15$ with none.

B. Experiment Two: Generalizing to Grocery Objects

Experiment Two demonstrated that our method generalized from seen basic shapes to unseen irregular groceries, achieving 95.0% success rate over both grocery sets. The YCB groceries, whilst geometrically accurate, have low mass and different friction properties compared to real groceries; yet, success rates for our PPO policy varied by only 1.4%, implying that performance was not sensitive to the weight or frictional properties of grasped objects. This is a consequence of the geometric caging approach, combined with limits on finger and palm reaction forces. The average successful grasp finger forces of 1.36N and palm force of 3.02N prevented observed damage to real groceries, except minor banana bruising. Lower force limits could be achieved, but reduce training quality.

The main limitation of our approach is that grasping invariance to weight and friction properties breaks down as masses increase. In practical terms, heavy objects cannot be grasped because the fingers are not stiff enough to lift them. Force limits and low friction exacerbate this issue, and the maximum weight of a smooth plastic ball that could be grasped was 550g using our best policy and $EI = 0.34$ fingers. With real groceries, most grasp failures occurred from the heaviest objects, like the banana bunch, slipping following an initial grip which formed a poor cage. Whereas, the YCB set included the smallest objects, and failures occurred from objects falling out from between the gripper fingers before they had closed.

We demonstrated that our method outperformed three baselines. The heuristic baseline had failed grasps due to naive feedback control, which only considered keeping forces in desired ranges, rather than accounting for state information and specific object geometry. Both MAT-based controllers had failures from poor action choices in out of distribution states, implying greater sensitivity to sim2real differences and unseen objects from the $12\times$ larger state vector. MAT generalized relatively well to our gripper, achieving 97.1% success rate on the 7 YCB food objects which were tested on and got 98.6% in the original paper. Our comparatively simpler state vector may aid generalization to other grippers, however the force limiting would be difficult to achieve without compliance and caging. The MAT baseline controller had the lowest stable 5N rate, 47.6%, as it did not learn to use the palm. This was because the minimum 1N palm force requirement was removed, to enable learning, and in line with the criteria used in the original paper.

Overall, Experiment Two demonstrated that the policy could generalize, supporting the use of noise during training coupled with compliance and caging. The flexible fingers acted like springs which repositioned objects into the center of grasp, whilst also balancing and smoothing out force signal irregularities. This meant that, despite training only on regular objects, reliable grasping was achieved on highly irregular items such as a net bag of tangerines, salad bag, and bunch of bananas. These objects are not only irregular in regard to shape but also dynamics, with changeable geometry and inertial properties during grasp. Compliance and caging aided generalization by handling uncertainty, which made objects appear less irregular to the policy. The gripper design was reliable, completing 2190 grasps in these experiments without mechanical faults.

VIII. Conclusion

We developed a grasping approach which combined reinforcement learning with force feedback, based on modeling compliant gripper fingers. We presented a gripper design and applied a model for the flexible gripper fingers, combining them to train in simulation a grasping controller with PPO. We demonstrated generalization to complex and challenging grocery objects whilst respecting force limits, with a 95.0% grasp success rate, and 80.1% of those grasps capable of resisting an additional 5N vertical disturbance. This exceeded performance of three baselines by 15.0% for success rate and 25.3% for resisting the 5N disturbance, despite up to 14.7% less in-grasp force. We evaluated our method on three finger rigidities, achieving 98.0% success rate on training objects and a 3.1% average sim2real gap. We showed that using the bending model and in-grasp sensors improved both learning and real world grasping performance. The main limitation of our approach was reduced performance for objects over 550g.

For future work, object localization and tolerance to clutter via camera sensors and additional actions will be considered.

References

[1] H. Sekkat *et al.*, "Review of reinforcement learning for robotic grasping: Analysis and recommendations," *Statistics, Optimization & Information Computing*, vol. 12, no. 2, pp. 571–601, Dec. 2023.

[2] S. Levine *et al.*, "Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection," *International Journal of Robotics Research*, vol. 37, no. 4-5, pp. 421–436, 2018.

[3] J. Mahler *et al.*, "Learning ambidextrous robot grasping policies," *Science Robotics*, vol. 4, no. 26, 2019.

[4] D. Kalashnikov *et al.*, "QT-Opt: Scalable deep reinforcement learning for vision-based robotic manipulation," *arXiv:1806.10293*, 2018.

[5] D. Morrison, P. Corke, and J. Leitner, "Learning robust, real-time, reactive robotic grasping," *The International Journal of Robotics Research*, vol. 39, no. 2-3, pp. 183–201, 2020.

[6] E. Ardissone, S. Ulrich, and A. Kirchheim, "Design and evaluation of an automatic decision system for gripper selection in order picking," *Logistics Journal: Proceedings*, vol. 2023, no. 1, 2023.

[7] F. Angelini *et al.*, "SoftHandler: An integrated soft robotic system for handling heterogeneous objects," *IEEE Robotics and Automation Magazine*, vol. 27, no. 3, pp. 55–72, 2020.

[8] H. Mnyusiwalla *et al.*, "A bin-picking benchmark for systematic evaluation of robotic pick-and-place systems," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 1389–1396, 2020.

[9] W. Friedl and M. A. Roa, "CLASH — a compliant sensorized hand for handling delicate objects," *Frontiers in Robotics and AI*, vol. 6, no. January, pp. 1–15, 2020.

[10] J. Hernandez *et al.*, "Current designs of robotic arm grippers: A comprehensive systematic review," *Robotics*, vol. 12, no. 1, 2023.

[11] Y. Chebotar *et al.*, "Generalizing regrasping with supervised policy learning," in *2016 International Symposium on Experimental Robotics*. Springer, 2017, pp. 622–632.

[12] H. Merzić *et al.*, "Leveraging contact forces for learning to grasp," in *2019 International Conference on Robotics and Automation (ICRA)*, 2019, pp. 3615–3621.

[13] J. Qu *et al.*, "Recent progress in advanced tactile sensing technologies for soft grippers," *Advanced Functional Materials*, vol. 33, no. 41, p. 2306249, 2023.

[14] A. Lobbezoo, Y. Qian, and H.-J. Kwon, "Reinforcement learning for pick and place operations in robotics: A survey," *Robotics*, vol. 10, no. 3, 2021.

[15] G. Mengaldo *et al.*, "A concise guide to modelling the physics of embodied intelligence in soft robotics," *Nature Reviews Physics*, vol. 4, no. 9, pp. 595–610, 2022.

[16] H. Choi *et al.*, "On the use of simulation in robotics: Opportunities, challenges, and suggestions for moving forward," *Proceedings of the National Academy of Sciences*, vol. 118, no. 1, p. e1907856118, 2021.

[17] L. Beddow, H. Wurdemann, and D. Kanoulas, "A caging inspired gripper using flexible fingers and a movable palm," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021, pp. 7195–7200.

[18] A. Zeng *et al.*, "Learning synergies between pushing and grasping with self-supervised deep reinforcement learning," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018, pp. 4238–4245.

[19] Z. Chen, Z. Jia, M. Lin, and S. Jian, "Towards generalization and data efficient learning of deep robotic grasping," in *IEEE Conference on Industrial Electronics and Applications (ICIEA)*, 2022, pp. 804–809.

[20] S. Song, A. Zeng, J. Lee, and T. Funkhouser, "Grasping in the wild: Learning 6DoF closed-loop grasping from low-cost demonstrations," *Robotics and Automation Letters*, 2020.

[21] T. Kim *et al.*, "Acceleration of actor-critic deep reinforcement learning for visual grasping by state representation learning based on a preprocessed input image," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021, pp. 198–205.

[22] F. Liu *et al.*, "Hybrid robotic grasping with a soft multimodal gripper and a deep multistage learning scheme," *IEEE Transactions on Robotics*, vol. 39, no. 3, pp. 2379–2399, 2023.

[23] Y. Deng *et al.*, "Deep reinforcement learning for robotic pushing and picking in cluttered environment," in *2019 IEEE/RSJ Conference on Intelligent Robots and Systems (IROS)*, 2019, pp. 619–626.

[24] B. Wu, I. Akinola, J. Varley, and P. K. Allen, "MAT: Multi-fingered adaptive tactile grasping via deep reinforcement learning," in *Proceedings of the Conference on Robot Learning*, vol. 100, 2020, pp. 142–161.

[25] B. Wu, I. Akinola, and P. K. Allen, "Pixel-attentive policy gradient for multi-fingered grasping in cluttered scenes," in *IEEE/RSJ Conference on Intelligent Robots and Systems (IROS)*, 2019, pp. 1789–1796.

[26] V. Kumar *et al.*, "Contextual reinforcement learning of visuo-tactile multi-fingered grasping policies," in *NeurIPS: Robot learning workshop*, 2019.

[27] H. Liang *et al.*, "Multifingered grasping based on multimodal reinforcement learning," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 1174–1181, 2022.

[28] C.-H. Liu *et al.*, "Optimal design of a motor-driven three-finger soft robotic gripper," *IEEE/ASME Transactions on Mechatronics*, vol. 25, no. 4, pp. 1830–1840, 2020.

[29] R. Adam Bilodeau, E. L. White, and R. K. Kramer, "Monolithic fabrication of sensors and actuators in a soft robotic gripper," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2015, pp. 2324–2329.

[30] S. Cheng *et al.*, "Stick-on large-strain sensors for soft robots," *Advanced Materials Interfaces*, vol. 6, no. 20, p. 1900985, 2019.

[31] R. Liu, S. Wang, H. Yang, and C. Shi, "Highly stretchable strain sensor with spiral fiber for curvature sensing of a soft pneumatic gripper," *IEEE Sensors Journal*, vol. 21, no. 21, pp. 23 880–23 888, 2021.

[32] L. Kuang, Y. Lou, and S. Song, "Design and fabrication of a novel force sensor for robot grippers," *IEEE Sensors Journal*, vol. 18, no. 4, pp. 1410–1418, 2018.

[33] S. Hogreve and K. Tracht, "Design and implementation of multiaxial force sensing gripper fingers," *Production Engineering*, vol. 8, no. 6, pp. 765–772, 2014.

[34] I. Huang *et al.*, "DefGraspSim: Physics-based simulation of grasp outcomes for 3D deformable objects," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 6274–6281, 2022.

[35] S. Navaee and R. E. Elling, "Equilibrium configurations of cantilever beams subjected to inclined end loads," *Journal of Applied Mechanics, Transactions ASME*, vol. 60, no. 2, p. 564, 1993.

[36] L. L. Howell and A. Midha, "Parametric deflection approximations for end-loaded, large-deflection beams in compliant mechanisms," *Journal of Mechanical Design*, vol. 117, no. 1, pp. 156–165, 03 1995.

[37] H. Su, "A pseudorigid-body 3R model for determining large deflection of cantilever beams subject to tip loads," *Journal of Mechanisms and Robotics*, vol. 1, no. 2, 01 2009, 021008.

[38] Vedant and J. T. Allison, "Pseudo-rigid-body dynamic models for design of compliant members," *Journal of Mechanical Design*, vol. 142, no. 3, p. 031116, 2020.

[39] V. K. Venkiteswaran, J. Sikorski, and S. Misra, "Shape and contact force estimation of continuum manipulators using pseudo rigid body models," *Mechanism and Machine Theory*, vol. 139, pp. 34–45, 2019.

[40] M. Khoshnam and R. V. Patel, "A pseudo-rigid-body 3R model for a steerable ablation catheter," in *2013 IEEE International Conference on Robotics and Automation*, 2013, pp. 4427–4432.

[41] R. J. Roesthuis and S. Misra, "Steering of multisegment continuum manipulators using rigid-link modeling and FBG-based shape sensing," *IEEE Transactions on Robotics*, vol. 32, no. 2, pp. 372–382, 2016.

[42] K. E. Bisshopp, "Approximations for large deflection of a cantilever beam," *Quarterly of Applied Mathematics*, vol. 30, no. January, pp. 521–526, 1973.

[43] D. Peckner and I. Bernstein, *Handbook of Stainless Steels*. New York: McGraw-Hill Book Company, 1977.

[44] J. Schulman *et al.*, "Proximal policy optimization algorithms," *arXiv:1707.06347*, 2017.

[45] J. Schulman *et al.*, "High-dimensional continuous control using generalized advantage estimation," *arXiv:1506.02438*, 2015.

[46] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *International Conference on Learning Representations, ICLR*, 2015.

[47] B. Calli *et al.*, "The YCB object and model set: Towards common benchmarks for manipulation research," in *2015 International Conference on Advanced Robotics (ICAR)*, 2015, pp. 510–517.

Luke Beddow (Student member, IEEE) received the M.Eng. degree in mechanical engineering from the University of Bath in 2020. He is currently working towards the PhD degree jointly in computer science and mechanical engineering with University College London in the Robot Perception and Learning lab. His research interests include robotic grasping, mechanism design, and reinforcement learning.

Helge Wurdemann (Member, IEEE) is Professor of Robotics at University College London leading the Soft Haptics and Robotics lab in the Department of Mechanical Engineering. He has been selected Turing Fellow at the The Alan Turing Institute. Prior, he received a degree (Dipl.-Ing.) in electrical engineering from the Leibniz University of Hanover, and a PhD in Robotics from King's College London in 2012. Helge has authored over 100 articles, published in high-impact journals.

Dimitrios Kanoulas (Member, IEEE) received the PhD degree from Northeastern University, Boston. He was a Postdoctoral Researcher at the Italian Institute of Technology for five years. He is currently a Professor in Robotics and AI with the Department of Computer Science, University College London, and a UKRI Future Leaders Fellow. He has published over 70 research papers in high-impact robotic journals and conferences. His research interests include robot perception, planning, and learning.