

How Firm Should a Grasp Be?

Matthew Beveridge
Columbia University
beveridge@cs.columbia.edu

Shree K. Nayar
Columbia University
nayar@cs.columbia.edu

Abstract: An ideal robot grasp is firm enough to securely handle an object, yet gentle enough to avoid damaging it. Achieving this balance requires knowledge of the object’s material properties, such as its mass, elasticity, and surface friction. These properties, however, are seldom precisely known *a priori*. In this work, we propose a visuotactile approach to estimating material properties in real time, during the process of grasping. Our method uses these estimated properties to determine the minimum grasp force required to handle the object. We contribute a new dataset of real-world objects (fruits and vegetables) with measured physical properties (shape, mass, elasticity, and friction), which we use to construct our force estimation model via simulations. We experimentally validate our approach to grasp force control using a robot with a parallel-jaw gripper. We demonstrate our system’s ability to gently grasp a wide variety of objects, in each case adapting to their unique physical properties.

Keywords: Material Recognition, Deformable Object Grasping, Tactile Sensing, Force Control

1 Gentle Grasping Requires Material Perception

Given an object and a desired interaction with it, the ideal grasp should exert no more force on the object than necessary to complete the task. While the notion of damage is often considered in catastrophic terms (*e.g.*, a glass shattering), it can also be more subtle. For example, with excessive force, a fruit may bruise, a cardboard box might dent, or a plastic toy could crack or get scratched. In addition, there is a temporal element to damage – repeated interactions without care for an object’s fragility can cause cumulative wear over time. In this work, we are interested in the *gentle grasping* of objects, where the goal is to minimize the potential for damage during a grasp.

Since the material properties of an object are seldom precisely known *a priori*, how do we, humans, determine how firm its grasp should be? Dynamic grasp adaptation is intuitive to us [1, 2] – we use our vision to obtain priors on the object’s material composition [3–5], and rely on tactile feedback to refine these priors while grasping it [6, 7]. Since some of the material cues we need to understand an object’s physics cannot be known from vision alone, we estimate them via touch during the nascent stage of a grasp. We use these estimates to continuously adjust the forces we apply to the object and are thus able to gracefully handle it without conscious effort.

For a robot to gently grasp an object, it must either know the object’s properties beforehand or be able to infer them during the grasp. Unless the object is precisely engineered or measured in detail, the best the robot has to work with is a rough approximation of its properties. One method to refine this approximation is for the robot to interact with the object before attempting a grasp. That is, the robot “probes” the object with a sequence of exploratory actions to estimate, for instance, its hardness [8, 9]. Such an approach is time-consuming and hence not suitable for applications where speed is critical. Instead, we propose to leverage the dynamic, ephemeral tactile signals that are sensed *during a grasp* to estimate the object’s properties in real time. The robot is therefore able to adjust the force it applies to the object on-the-fly, *in situ*.

In a typical robotic task, there are four initial phases, as illustrated in Fig. 1. The first is *observe*, where, from a distance, the robot can visually identify the class of the object, its shape, and its location. During the *approach* phase, the robot positions itself at a suitable grasp location, determined from the object’s geometry and pose. Next, we have *grip*, where the gripper squeezes the object with a specified force to take hold of it. Finally, the robot *lifts* the object and executes a goal task, such as moving it to a new location or mating it with another object. In our work, we focus on the *observe*, *grip*, and *lift* phases and assume that for *approach* a reasonable grasp pose is provided by

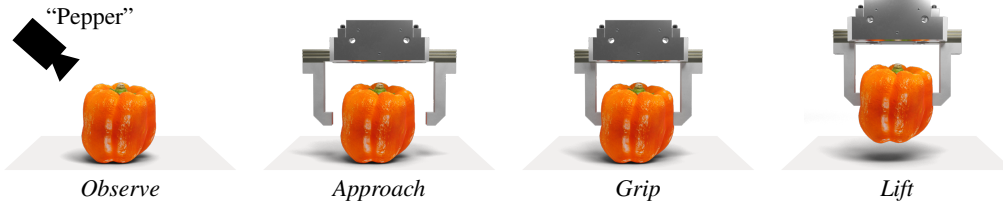


Figure 1. **Phases of a Grasp.** We consider a grasp to comprise four stages: observe, approach, grip, and lift. Our work exploits the short-lived (fraction of a second) haptic signals during the grip and lift phases to refine class-level material properties (from observation) to obtain instance-level estimates. During this refinement, we continuously adjust the applied force to ensure it is just above the minimum force needed to handle the object.

an existing planner [10, 11]. The *observe* phase reveals the class of the object (“pepper” in Fig. 1), giving us a prior on its properties (mass, elasticity, friction, *etc.*). During the *grip* and *lift* phases, we use haptic feedback to continuously refine these priors and control the force applied to the object.

While haptic signals encode an object’s material properties, different combinations of material parameters could lead to the same haptic signals. Therefore, material estimation can be viewed as a challenging inverse problem. To develop our solution to this inverse problem, we have chosen the domain of fruits and vegetables. This is an interesting domain as two samples, even when of the same class, can be expected to differ in their properties (*e.g.*, a raw and a ripe avocado). We first collected a dataset, called *Squash*, that includes material measurements and 3D scans of 270 samples across 52 different classes of fruits and vegetables.¹ Using *Squash*, we derived class-level property distributions, which reduce the search space for instance-level properties.

Using *Squash*, we simulated numerous object grasps, with augmentations of material properties to span real-world variability. We trained a sequence model on the evolution of the simulated haptic data (*i.e.*, tactile images and load cell measurements) to distill instance-level material properties from class distributions. We have integrated our visuotactile model for force control into a robot’s operating system, where it runs at 250 Hz. We used a 6 DoF arm and a parallel-jaw gripper with GelSight tactile sensors to validate our approach.² Our system achieved more than a $2\times$ reduction in excess applied force for a successful grasp compared to non-adaptive baselines.

2 Related Work

Class-Level Material Recognition from a Distance. For grasping, it is beneficial to recognize an object’s material class (*e.g.*, peach) passively during the *observation* stage. However, precise class-level visual recognition (*i.e.*, peaches vs. nectarines) requires elaborate optical measurements, such as the reflectance distribution function (BRDF) [12–14], which is hard in dynamic, cluttered robotics settings. Hence, class recognition is most often done from appearance alone [15–19], using filter banks, statistical analysis, and deep learning. These approaches do not directly relate appearance to material properties, which has been the goal of more recent work [20–22]. In our work, we rely on appearance to identify class-level material properties during *observation*, while recognizing that additional sensing modalities are needed to resolve instance-level variation within a class.

Instance-Level Material Recognition through Excitation. During the *grip* and *lift* stages, haptic feedback provides material-related signals for the specific object of interest (*e.g.*, *this* peach has *this* hardness). In the general context of measuring physical properties, exciting objects with ultrasound and physical impact have been shown to reveal useful material cues [23–25]. In the realm of robotics, vision-based tactile sensors have been used to provide detailed insights related to contact mechanics. For example, GelSight [26–28] and DIGIT [29] have been used for hardness estimation [30] and shear/slip detection [31]. Beyond optical sensors, piezoresistive [32, 33] and magnetic tactile skins [34], which measure pressure distributions, have been integrated into grippers for fine-grained visuotactile manipulation [35, 36]. Cross-modal methods have also been developed to relate visual appearance to tactile properties [21, 37–40]. In our work, we use GelSight sensors to image contact

¹The dataset is available on the [Squash webpage](#).

²Videos are available on the [project webpage](#).

geometry, shear, and slip during a grasp. Our framework for material-aware grasping, however, is applicable across a variety of tactile sensing modalities.

Approaches to Gentle Grasping. A classical approach to avoiding object damage is to regulate contact force via impedance control [41], or to adjust grasp force reactively using tactile feedback [42–44], *e.g.*, upon detecting incipient slip [45, 46]. For this purpose, tactile sensors have also been integrated into soft robotic hands due to their passive compliance [47, 48]. Purely vision-based methods [49–55] have also demonstrated successful deformation-aware grasping, but are shown to be more effective when fused with tactile sensing [56, 57]. More recently, human and robot demonstrations have enabled end-to-end gentle grasping of objects [8, 9, 36, 58, 59], and diffusion-based policies conditioned on tactile signals have been used for force-aware manipulation of fragile objects [60]. Nevertheless, these methods are not real-time, do not expose inferences about the object’s material, and only yield predictions of an “acceptable” grasp force based on heuristics.

3D Object Datasets. Several large-scale 3D object datasets have been developed for robotic manipulation [61–65] and broader 3D vision tasks [66–68]. These datasets provide geometry and appearance but do not include measured material properties. The recent work by Cao and Kalogerakis [69] uses annotated simulated objects with material parameters, where the parameter values are assigned using material-category labels rather than measurements made on individual instances. For deformable objects, smaller datasets exist that include detailed annotations of the deformation process through finite element analysis [49, 70]. One of our contributions is *Squash*, a dataset that pairs 3D scans of fruits and vegetables with per-instance measurements of material properties (elasticity, mass, and surface friction). It is the first 3D object dataset specifically designed for deformable object simulation with measured properties. Beyond its utility in robotics, we expect *Squash* to be a valuable resource for simulation and rendering in computer graphics and animation.

3 The Physics of Grasping a Material

If we know the material properties of an object, then what minimal grasp force should we use to lift it? Naïvely, the force F required to suspend a rigid body of mass m at equilibrium is $F = mg$, *i.e.*, Newton’s second law. If we assume that the force due to friction is independent of the mass and the contact area, we would get the familiar modification: $\mu F = mg$, where μ is the coefficient of friction. Coulomb observed that the effect of friction is primarily due to the interlocking of surface asperities, which he assumed to be infinitesimally small. This, however, is an approximation. Bowden and Tabor [71] noted that friction capacity can vary under pressure, thereby acting as an additional load, especially when the interacting bodies are soft. The effect of this additional load is visualized in Fig. 2, where we see that soft objects require a lower grasp force than rigid ones, since they deform and create a larger contact area under the same load. Formally, this yields a new equilibrium:

$$\mu F + \tau A_{\text{real}} = mg, \quad (1)$$

where τ is the interfacial shear strength, and A_{real} is the contact area over which surface asperities interact. For rigid bodies, A_{real} is considerably smaller than the apparent contact region A_{app} (on the macroscale), as only the peaks of the asperities interlock [72].

If the object or the gripper is soft, we can say $A_{\text{real}} \approx A_{\text{app}}$, since the elastic deformation fills the cavities between the summits of surface asperities. In our context, A_{app} can be directly observed through tactile sensing. Even so, F and A_{app} are directly coupled, and without a mapping between

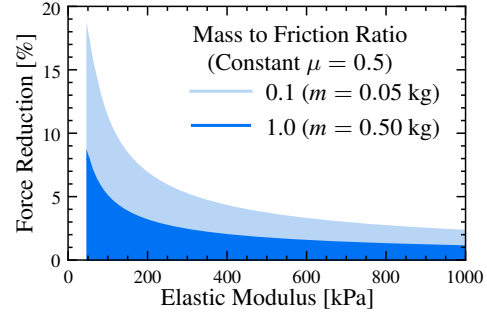


Figure 2. Soft Objects Require Less Grasp Force. For a soft object, the contact area A grows more rapidly with applied force F than it does for a hard object. As A grows, an additional load, τA , where τ is the object’s material shear strength, supplements F to resist the effect of gravity (Eq. (1)). The theoretical reduction in minimum-required grasp force due to this additional load is shown here as a function of the elastic modulus. The object and its shear strength τ are kept constant in the plot, but the ratio of mass m to friction μ is varied. The greatest benefit (*i.e.*, the possible reduction in grasp force) is for light, pliable objects.

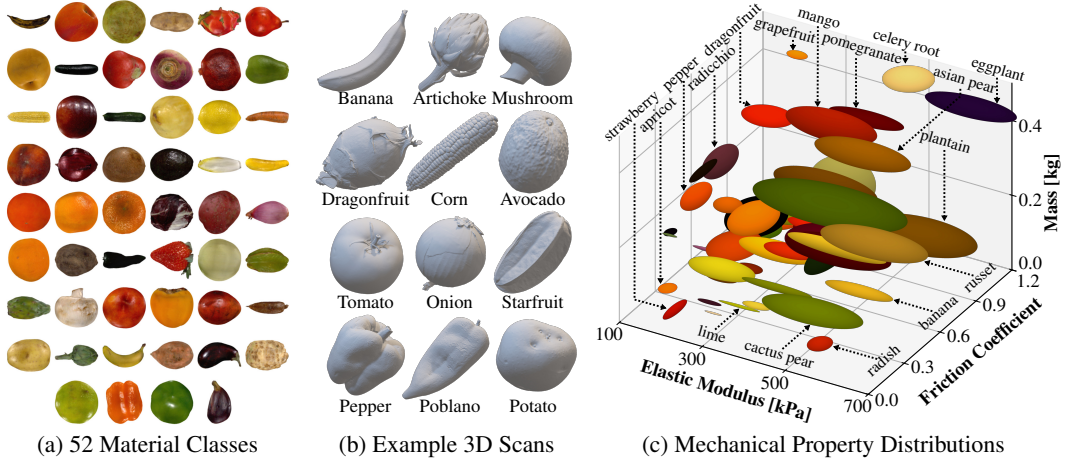


Figure 3. **The Squash Dataset.** A dataset of real-world 3D object meshes and material property measurements. (a) *Squash* is composed of 270 instances across 52 classes of produce (fruits and vegetables). (b) We scanned each object using structured light to measure its geometry and texture at high resolution. (c) We additionally measured the material properties of each object (elastic modulus, mass, volume, and surface friction), where each ellipsoid represents the distribution of the properties within a class. *Squash* is a first-of-its-kind dataset that captures the natural variability of object properties, which can be used for physically-accurate simulation.

them, we must *search* for the force that results in the equilibrium given by Eq. (1). To *predict* this force of equilibrium, we must also predict how the apparent contact area will change with respect to the load force. Finite element analysis (FEA) can accurately estimate this contact area, assuming the material properties of the object and gripper are known, *a priori*. In our case, property estimates change during the course of a grasp, necessitating an expensive FEA reevaluation on each update, which would limit the rate at which we can adapt grasp force.

Instead, by using a Hertzian contact approximation [73, 74], A_{app} can be modeled using a closed-form expression. If we approximate the fingertips of a robotic gripper as near-planar, and an object as locally spherical, then applying a normal force F yields:

$$A_{\text{app}}(F) = \pi \left(\frac{3FR^*}{4E^*} \right)^{2/3}, \text{ given } \frac{1}{E^*} = \frac{1 - \nu_{\text{obj}}^2}{E_{\text{obj}}} + \frac{1 - \nu_{\text{grip}}^2}{E_{\text{grip}}} \text{ and } R^* = \frac{R_{\text{obj}} R_{\text{grip}}}{R_{\text{obj}} + R_{\text{grip}}}, \quad (2)$$

where E^* is the effective elastic modulus between the object and the gripper, R^* is the effective radius of curvature between the object and the gripper at the contact interface, and ν is the Poisson ratio. Given the mass m , elasticity E , surface friction μ , and interfacial shear strength τ of an object, we can find A_{app} using Eq. (2) and substitute in Eq. (1) to estimate the minimum force needed to lift the object. But, for an arbitrary object, we at best know a range of plausible values for (E, m, μ, τ) based on its class. Our goal, therefore, is to develop a learning-based approach to refine (E, m, μ, τ) through temporal changes in haptic feedback, which we then use to servo-control the applied force.

4 *Squash*: A Dataset of 3D Object Assets and their Physical Properties

To learn our visuotactile model for gentle grasping, we require a dataset of deformable objects with known physical properties. Deformable objects are interesting in two respects. First, they are more prone to damage due to a grasp that is too firm (*e.g.*, a peach will bruise). Second, as outlined in Sec. 3, the contact area between the gripper and the object changes as they both deform, thereby requiring continuous control of the minimum grasp force. We have developed *Squash* (Fig. 3), a comprehensive dataset of 3D scans and physical property measurements of various fruits and vegetables. *Squash* contains 270 instances of produce spread across 52 classes (capturing intraclass variability). All 52 classes are shown in Fig. 3a, and a few examples of the 3D scans are shown in Fig. 3b. Using the methods described below, we measured the elastic modulus, mass, volume, and surface friction of each instance. A visualization of the measured properties is shown in Fig. 3c, where each ellipsoid represents the distribution of the properties within a class.

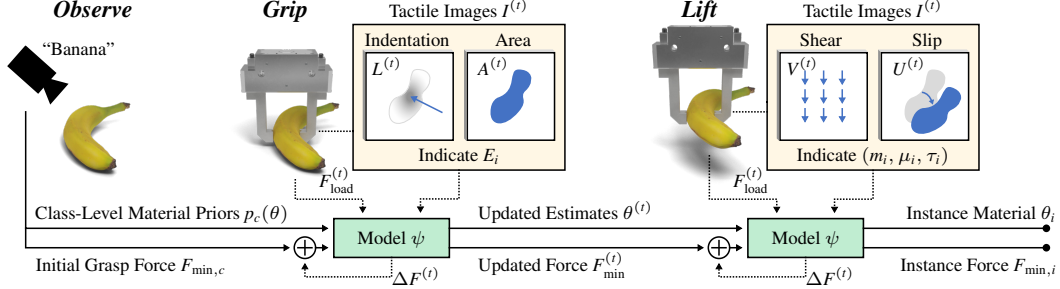


Figure 4. **Framework for Gentle Grasping.** From *observation*, we identify the object’s class to get a class-conditioned material prior and an initial minimum-required grasp force estimate. Upon first contact, our visuo-tactile likelihood model ψ (detailed in Fig. 5) begins to refine these priors into instance-level posteriors from temporal tactile feedback. The minimum-required grasp force is continuously reestimated in the process, and applied in real time. The result is a gentle grasp that adapts to the object instance, *in situ*.

Characterizing the Properties of an Object. For each object, we first measured its mass and Shore hardness using a durometer [75, 76]. From its 3D scan, we then computed its volume. Next, using a robot arm and a parallel-jaw gripper affixed with load cells and tactile fingertips (detailed in the supplemental material), we measured the object’s elasticity and friction capacity by performing a compression test and a tribological test, respectively. In the compression test, with the object at rest, we recorded the applied force F_{load} , contact area A , and indentation into the object L , while closing the gripper jaws on the object. We modeled the local surface of the object as quasi-spherical and the gripper jaws as parallel plates. Then, for any deformation of the object, the effective modulus E^* follows from Eq. (2): $E^* = \frac{3F}{4L} \sqrt{\frac{\pi}{A}}$, since from Hertzian mechanics, the effective curvature of the contact interface is $R^* = \frac{A}{\pi L}$. We have many estimates of effective modulus E^* (during closure), and we average them for our final estimate of E^* . We then recover the object’s elastic modulus E from E^* by separately measuring the elastic modulus of the tactile sensor. We report E as a function of the Poisson ratio ν , since we cannot directly observe the equatorial deformation of the object.

For the tribological test, we begin with the object lifted, then increase the gripper opening until slip is detected by the tactile sensor. Each trial yields one $(F_{\text{slip}}, A_{\text{slip}})$ pair, where F_{slip} is the applied force just before slip, and A_{slip} is the contact area at that onset. These measurements are related through Eq. (1): $\mu F_{\text{slip}} + \tau A_{\text{slip}} = mg$. We collected several trials per object, each at a different location on its surface, and solved the over-determined linear system in (μ, τ) by least squares. By sampling distinct locations on the object, we capture local deviations in surface curvature. This gives us different couplings between F_{slip} and A_{slip} , enabling the decoupling of μ and τ during estimation. Since *Squash* includes a large sampling of shapes, masses, elasticity parameters, and friction coefficients, we can use it to generate diverse grasp simulations for model training.

5 Material-Aware Robotic Grasping

Fig. 4 illustrates our closed-loop grasping framework for estimating material properties $\theta = (E, m, \mu, \tau)$ to determine the minimum necessary grasp force F_{min} for an object. The material prior distribution $p_c(\theta) \equiv p(\theta | c)$ for the object class c , identified during *observation*, yields an initial estimate of the minimum grasp force $F_{\text{min},c}$ via Eq. (1). The *visuotactile likelihood model* ψ then refines this class prior during contact with the object, yielding an online posterior belief over material properties at each time t , denoted as $\theta^{(t)}$. As the robot begins to grip the object, tactile images $I^{(t)}$ are used to update the elasticity estimate $E^{(t)}$ from temporal changes in indentation $L^{(t)}$ and contact area $A^{(t)}$ relative to the applied load $F_{\text{load}}^{(t)}$.³ As the robot begins to lift the object, shear $V^{(t)}$ and slip $U^{(t)}$ in the tactile images are used to update the estimated mass $m^{(t)}$ and friction parameters $\mu^{(t)}$ and $\tau^{(t)}$. The model ψ therefore continuously refines the material estimate $\theta^{(t)} = (E^{(t)}, m^{(t)}, \mu^{(t)}, \tau^{(t)})$, and the controller accordingly adjusts the applied grasp force $F_{\text{load}}^{(t)}$ to achieve the current minimum-force estimate $F_{\text{min}}^{(t)}$. As additional tactile observations arrive, the posterior approaches the instance-level material properties θ_i , yielding the desired object-specific force estimate $F_{\text{min},i}$ for gentle grasping.

³While our implementation uses GelSight tactile images, the model ψ can utilize other tactile sensor outputs, e.g., binary contact maps and pressure maps. For our purposes, we refer to all these outputs as tactile images.

Learning Material from Simulated Grasps. To develop our visuotactile likelihood model ψ , we first designed a simulator to generate high-fidelity tactile data for training. We built upon IPC [77–79] for collision physics, which has been shown to accurately model real-world grasps of deformable objects [80, 81], by modifying it to also model the adhesive effect of τ . From the empirical tribological tests in Sec. 4, we observe a Pearson correlation of 0.94 between the optimal grasp force and one predicted by our contact model in Eq. (1), upon which this simulator is built. We then synthesized 200k training grasp sequences using a parallel-jaw gripper and the object meshes in *Squash*, where object properties were sampled from the distributions in *Squash*. Each sequence contains temporal measurements of each jaw’s tactile image $I^{(t)}$ and applied normal load $F_{\text{load}}^{(t)}$. To make real-world estimation robust, we replicate our real-world load cell and GelSight sensor characteristics in simulation, add noise to their measurements, and supplement the *Squash* material properties by widely sampling simulated materials using Latin Hypercube sampling. We then train ψ , shaded green in Fig. 5, to encode these temporal haptic observations. At each timestep t , the tactile image $I^{(t)}$ is embedded by an image encoder, paired with $F_{\text{load}}^{(t)}$, and passed to a closed-form continuous-time neural network (CfC) [82] to produce a hidden state $h^{(t)}$. A CfC is chosen for its ability to model complex dynamics using ordinary differential equations, its ability to handle non-uniformly sampled robot data, and its inference speed. We trained ψ with a hybrid contrastive-regression loss over the simulated grasps. At inference, $h^{(t)}$ parameterizes the likelihood used to score candidate particles $\theta_j^{(t)}$ in a particle filter, producing an online posterior belief over material properties. Further simulation and training details are in the supplemental material.

Gentle Grasping from Temporal Observations. Using ψ and sequential observations $(I^{(0:t)}, F_{\text{load}}^{(0:t)})$, we update the posterior belief over material properties θ as:

$$p(\theta \mid I^{(0:t)}, F_{\text{load}}^{(0:t)}) \propto p_\psi(I^{(t)}, F_{\text{load}}^{(t)} \mid \theta) p(\theta \mid I^{(0:t-1)}, F_{\text{load}}^{(0:t-1)}), \quad (3)$$

where $p_\psi(\cdot \mid \theta)$ is the learned likelihood induced by ψ . We represent this belief with particles,

$$p(\theta \mid I^{(0:t)}, F_{\text{load}}^{(0:t)}) \approx \sum_j \alpha_j^{(t)} \delta(\theta - \theta_j^{(t)}), \quad \alpha_j^{(t)} \propto \alpha_j^{(t-1)} p_\psi(I^{(t)}, F_{\text{load}}^{(t)} \mid \theta_j^{(t)}), \quad (4)$$

where $\alpha_j^{(t)}$ is the weight of particle $\theta_j^{(t)}$. Since particles are resampled during filtering, both particle locations and weights are indexed by time. Each particle represents a candidate contact model and grasp feasibility condition [83, 84]. While lifting the object, feasibility reduces to whether the frictional support exceeds the object’s weight. The minimum required grasp force is therefore:

$$f(\theta) = \min_{F \geq 0} \{F : \mu F + \tau A(F; E) \geq mg\}, \quad (5)$$

following from Eq. (1), where $A(F; E)$ is given by Eq. (2). We compute a force hypothesis $F_{\text{min},j}^{(t)} = f(\theta_j^{(t)})$ for each particle, and the weighted set $\{F_{\text{min},j}^{(t)}, \alpha_j^{(t)}\}$ represents the current uncertainty in the required grasp force. As additional haptic observations arrive, posterior uncertainty over θ contracts, producing a tighter estimate of the minimum force needed to lift the object. Thus, ψ refines class-level material priors into instance-level estimates, enabling gentle, material-aware grasping, *in situ*.

6 Experiments

Evaluation via Simulations. We first compared our approach to open-loop baseline grasp force policies: *Naïve*, where the commanded grasp force is always 3 N (sufficient to lift the largest object in *Squash*) regardless of the object’s class, and *Class Prior*, where grasp force is determined using Eq. (1) from the class’s average material properties in *Squash*. We then compared our approach to adaptive slip-reactive force control methods, that increase force when incipient slip is

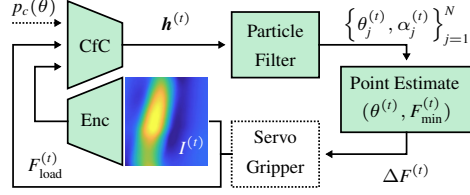


Figure 5. Visuotactile Model ψ (shaded green) for Material Estimation and Force Control. We start with the class prior $p_c(\theta)$ for an object’s material θ . Tactile images $I^{(t)}$ and the applied load $F_{\text{load}}^{(t)}$ are then encoded by the continuous-time sequence model (CfC) into the hidden state $h^{(t)}$. The particle filter uses this representation to update the weighted material belief $\{\theta_j^{(t)}, \alpha_j^{(t)}\}_{j=1}^N$. The resulting point estimate $\theta^{(t)}$ determines the current estimate of minimum-required grasp force $F_{\text{min}}^{(t)}$, which is used as the target force by the gripper servo-controller. The entire control loop (including ψ) runs at 250 Hz.

Table 1. **Efficacy of Material-Aware Robot Grasping.** We compare our material-aware approach for force control with open-loop and closed-loop baselines using 5k simulated objects with high material variability. “Naïve” uses a force of 3 N in each gripper jaw, sufficient to lift the largest object in *Squash*. “Class Prior” uses the class’s average material properties to determine the grasp force using Eq. (1). Direct force prediction estimates the grasp force without intermediate material estimation. Slip-reactive policies modulate grasp force to hover around the boundary where incipient slip is observed. In contrast, our approach refines instance-level material properties for fine-grained force control. The best results are in bold, and second best are underlined.

Grasp Force Controller	Force error			Control outcome		
	Bias [N]	MAE [N] ↓	WAPE ↓	Excess [N] ↓	Deficit [N] ↓	Success ↑
<i>Open-Loop Baseline Grasping Approaches (Non-Adaptive)</i>						
Naïve constant force (3 N)	1.97 [1.67, 2.22]	2.18 [2.02, 2.33]	2.11 [1.56, 2.95]	2.07 [1.87, 2.26]	0.10 [0.01, 0.23]	1.00 [0.99, 1.00]
Class prior	-0.26 [-0.57, 0.10]	0.70 [0.50, 0.95]	0.68 [0.59, 0.77]	0.22 [0.14, 0.31]	0.48 [0.27, 0.74]	0.89 [0.77, 1.00]
<i>Closed-Loop Baseline Grasping Approaches (Adaptive)^a</i>						
Direct force predictor (learned)	0.29 [-0.05, 0.60]	0.96 [0.75, 1.19]	0.93 [0.74, 1.21]	0.62 [0.46, 0.80]	0.33 [0.14, 0.59]	0.96 [0.90, 1.00]
Slip-reactive (heuristic, shear)	0.77 [0.52, 0.98]	1.12 [0.98, 1.29]	1.09 [0.86, 1.43]	0.95 [0.82, 1.08]	<u>0.18</u> [0.05, 0.34]	0.96 [0.91, 1.00]
Slip-reactive (heuristic, contact area)	-0.52 [-0.82, -0.26]	0.71 [0.48, 0.99]	0.69 [0.60, 0.75]	0.10 [0.06, 0.17]	0.61 [0.37, 0.90]	0.93 [0.87, 1.00]
Slip-reactive (learned, GBM [86])	-0.51 [-0.79, -0.24]	0.72 [0.49, 1.00]	0.70 [0.62, 0.76]	0.10 [0.07, 0.13]	0.62 [0.38, 0.91]	0.93 [0.86, 0.99]
Slip-reactive (learned, images)	0.16 [-0.29, 0.51]	0.88 [0.69, 1.11]	0.85 [0.73, 1.01]	0.47 [0.37, 0.56]	0.41 [0.19, 0.68]	0.97 [0.93, 1.00]
<i>Material-Aware Closed-Loop Grasping Approaches (Adaptive), with Ablations^b</i>						
Ours	-0.05 [-0.44, 0.24]	0.51 [0.19, 0.84]	0.52 [0.43, 0.60]	<u>0.21</u> [0.16, 0.26]	0.33 [0.09, 0.68]	0.95 [0.90, 0.99]
↳ no prior	-0.14 [-0.42, 0.07]	<u>0.61</u> [0.43, 0.83]	<u>0.59</u> [0.49, 0.68]	0.23 [0.17, 0.29]	0.38 [0.19, 0.62]	0.94 [0.84, 1.00]
↳ wrong (adversarial) prior	-0.20 [-0.46, 0.03]	0.62 [0.47, 0.86]	<u>0.59</u> [0.50, 0.67]	<u>0.21</u> [0.14, 0.30]	0.40 [0.21, 0.64]	0.93 [0.82, 0.97]
↳ Coulomb-only friction	-0.16 [-0.41, 0.08]	0.62 [0.45, 0.84]	0.61 [0.52, 0.69]	0.24 [0.19, 0.30]	0.38 [0.21, 0.68]	0.94 [0.88, 0.98]
Ours + safety margin ^c	0.22 [-0.14, 0.59]	1.00 [0.77, 1.26]	0.97 [0.78, 1.24]	0.61 [0.40, 0.86]	0.39 [0.20, 0.61]	<u>0.99</u> [0.96, 1.00]

^a PR-AUC of slip detection methods: shear (0.05), contact area (0.16), GBM (0.23), images (0.80). Higher is better, random is 0.04.

^b Ablations: “no prior” removes prior conditioning; “wrong prior” uses a mismatched prior at test time; “Coulomb-only” removes the τA term from the contact model used to determine required force from estimated material properties.

^c Commands $\gamma \cdot Q_p(F_{\text{target}})$ over the posterior; ($\gamma = 1.25, p = 0.5$) selected on validation objects.

detected [85], and to a direct force predictor that estimates grasp force without estimating material properties. Using the simulator described in Sec. 5, all learned estimators are trained and validated on 200k simulated pick-ups of 85% of the *Squash* objects, and evaluated on 5k simulations of the remaining 15%. Both sets use randomly augmented 3D meshes and material properties sampled from the *Squash* dataset, where we procedurally augment meshes to create realistic variations in shape. The ground truth ideal grasp force is determined via a “drop test” where applied force (at the evaluated grasp pose) is decreased until gross slip is observed. The results are summarized in Tab. 1. Our material-adaptive approach outperforms baselines through accurate estimation of the object’s material properties (Pearson correlations: $r_E=0.65$, $r_m=0.88$, $r_\mu=0.28$, $r_\tau=0.57$), demonstrating its ability to adapt to the specific material properties of each object and maintain a high success rate. Notably, our approach produces an unbiased force estimation, closest to the ideal value of any method, showing closed-loop material estimation is beneficial for gentle grasping. We additionally ablate the class prior and τA term in the Bowden-Tabor contact model (Eq. (1)) to isolate their contributions. We observe small performance degradations in each case, indicating that material priors are beneficial but not strictly necessary, and that the τA term improves force accuracy to the extent outlined in Sec. 3. The limited degradation from prior ablation also suggests our method would generalize well beyond the domain of fruits and vegetables.

Grasping Real Fruits and Vegetables. We next assembled a test set of 29 real fruits and vegetables. Their material properties were measured (using the methods in Sec. 4) as a ground truth reference. We then performed 3 grasping trials per object with our material-aware approach, each utilizing a different grasp pose, using the robot hardware shown in Fig. 6a. After completing each grasping trial, at the executed grasp pose, the applied force is reduced until gross slip is observed to empirically determine the minimum required grasp force for the object. In Figs. 6b and 6c, we show the evolution of the applied grasp force and estimated material properties over the duration of grasping a nectarine. As the robot progresses through the *grip* phase, the estimate of elasticity is refined. During the *lift* phase, mass and friction parameters are updated. The force and material estimates converged to stable values quickly over the course of the grasp, enabling the robot to apply a force that never significantly exceeds the optimal force for the nectarine. Fig. 6d compares the grasp forces of our method with those of the *Naïve* and *Class Prior* baselines, for a wide variety of fruits and vegetables. As seen in Fig. 6e, the distribution of excess forces for our approach is more tightly concentrated around zero, compared to the baselines. Finally, in Figs. 6f and 6g, we show an application of our

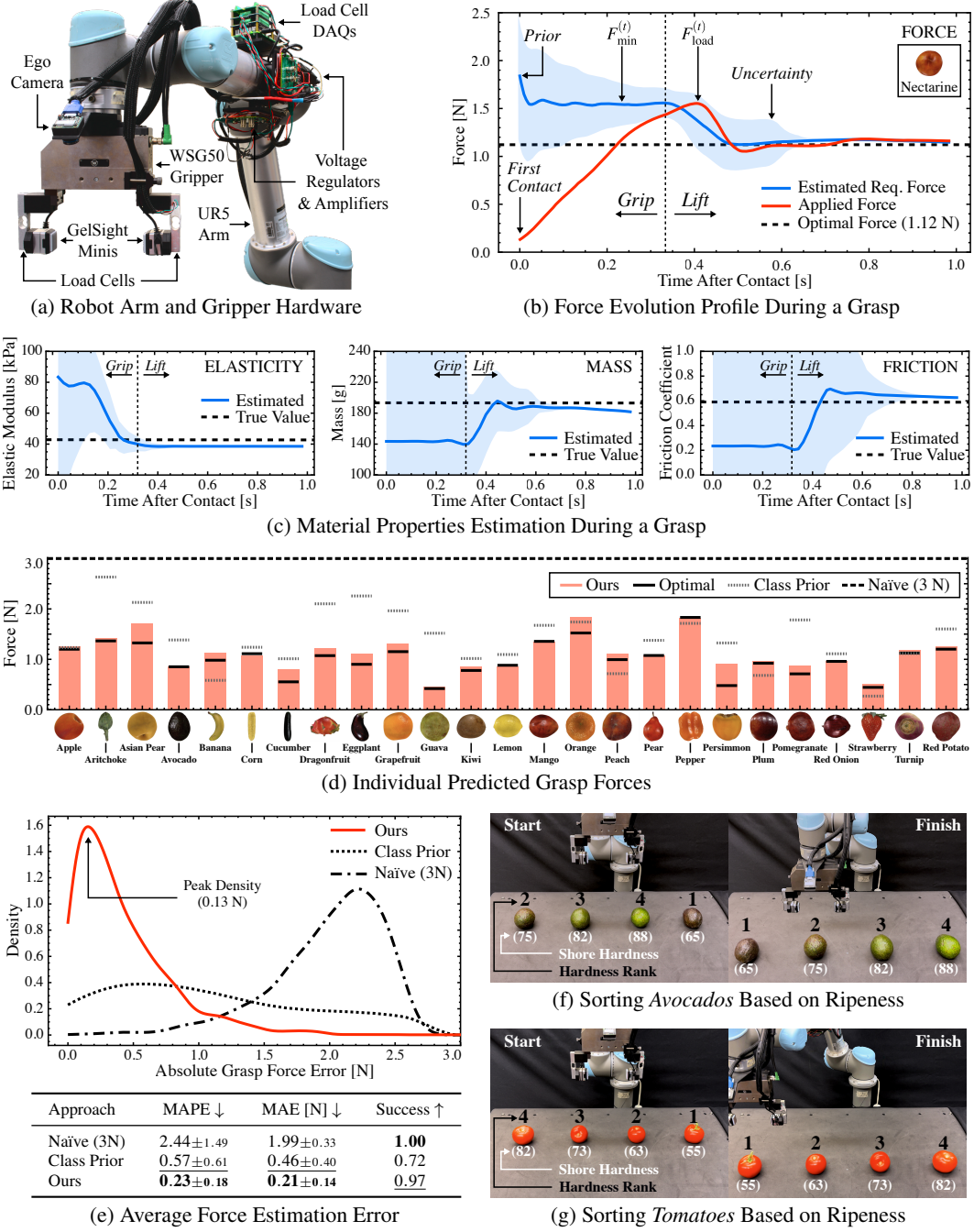


Figure 6. **Experimental Results.** (a) Robot hardware setup for real-world grasping experiments. Our approach is powered by GelSight Mini tactile sensors paired with bending beam load cells, but is not limited to this configuration. (b-c) Evolution of the applied grasp force and estimated material properties during a trial. The values converge quickly, and, as a result, we gently grasp the object in real time with a force close to the optimal force. (d) The applied grasp force by our approach for various real fruits and vegetables. We consistently apply at or slightly above the optimal force. (e) Performance of our approach over all 29 tested real fruits and vegetables, where we show accurate force control while maintaining a high grasp success rate. In the kernel density plot, our approach shows a tighter concentration around zero, indicating consistent performance that adapts to individual object properties. (f-g) Sorting fruits in real time based on their estimated elasticity, which is correlated with ripeness. This is one potential application of material-aware grasping beyond force control.

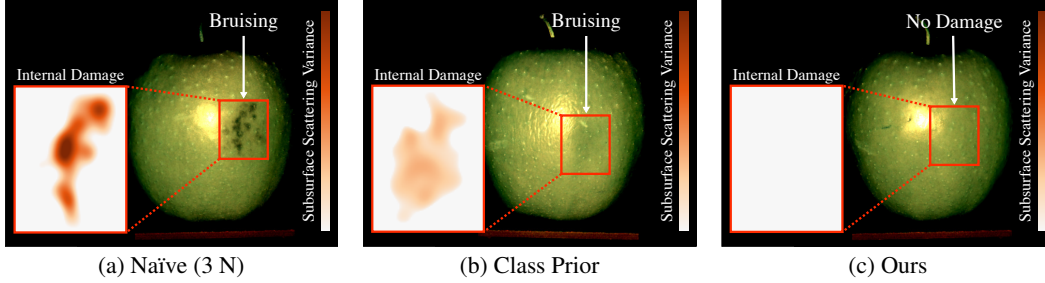


Figure 7. **Internal Apple Damage per Grasp Method.** All images captured using fast separation [87] immediately after force application, revealing subsurface anomalies invisible to the human eye. High variation in subsurface scattering indicates irregular water content under the apple’s skin, which is correlated with bruising. The (a) naïve and (b) class prior approaches both leave signs of damage, while (c) our approach does not.

approach to a real-world problem – sorting fruits based on their level of ripeness. As ripeness is correlated with elasticity in climacteric fruits, by estimating the elasticity of each fruit during its grasp, the robot can sort them based on their ripeness in real time.

Qualitative Damage Evaluation. Since applying excess force increases risk of damaging the object being grasped, we have thus far used excess force as a quantitative proxy for damage. However, using structured light imaging [87], we can reveal internal bruising from each grasping method (Fig. 7). By projecting high-frequency illumination, we can image internal variations in the subsurface scattering of an apple immediately after each grasp was completed. Further, in Fig. 8, we can see surface damage to the banana 1 day after being grasped. In both cases, only our approach shows no visible damage.

7 Limitations and Directions for Future Work

We have taken a first step towards robot grasping that minimizes the force used to grasp an object. In our work, we focused on the task of controlling force given a suitable grasp pose for a parallel-jaw gripper.

To minimize the potential for damage, we would also need to address the problem of choosing grasp poses that do not require large forces (*e.g.*, a banana picked up by one end). While we have used a parallel-jaw gripper, further analysis of the physics of grasping would be needed to extend our work to more complex end-effectors, such as dextrous hands that can cradle objects. Additionally, the performance of our method is tied to that of the tactile sensor we use. Current sensors trade off spatial and temporal resolution. While we addressed this problem in our work by using interpolation, to improve performance we will need sensors with both high update rate and spatial resolution. Furthermore, we have focused on the domain of fruits and vegetables, for which we could measure class priors. For objects without class priors, our method may still be effective, but this would need to be validated via extensive experiments. Finally, we have assumed that the object’s material properties are more or less constant over its surface. We plan to extend our work to the harder problem of grasping objects with spatially varying properties (*e.g.*, a toy stuffed animal).

8 Conclusion

We have introduced material-aware robotic grasping: a visuotactile approach that estimates instance-level material properties, online, during contact with an object. This allows us to drive closed-loop force control towards the theoretical minimum grasp force. We created *Squash*, a diverse dataset of 3D object models with measured material properties. Our visuotactile model, trained using *Squash*, recovers elasticity, mass, and friction parameters from tactile feedback. In simulation and on real hardware, our method consistently achieved gentle, non-damaging grasps. In addition, the *in situ* material estimation of our method opens avenues for object sorting, grasp-pose optimization, dexterous manipulation, or any task where knowing *how* an object deforms is of value.

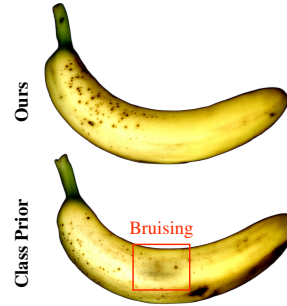


Figure 8. **Grasp Damage After 1 Day.** Contrast-enhanced images of a banana 1 day after grasp force application. (Top) The grasp force predicted by our approach leaves no visible damage. (Bottom) The class prior approach overexerts, leaving a bruise.

Acknowledgments

This work was supported in part by the Office of Naval Research (ONR) award N00014-24-1-2155 and in part by the National Science Foundation (NSF) and Center for Smart Streetscapes (CS3) under NSF Cooperative Agreement No. EEC-2133516. The authors are grateful to Behzad Kamgar-Parsi at ONR for his support. The authors thank Dhruv Yalamanchi and Isabel Tu for help with collecting the *Squash* dataset, Joaquin Palacios for discussions on robot control, and Silvia Sellán for visualization support.

References

- [1] V. C. Paulun, U. Kleinholdermann, K. R. Gegenfurtner, J. B. J. Smeets, and E. Brenner. Center or side: Biases in selecting grasp points on small bars. *Experimental Brain Research*, 232(7): 2061–2072, July 2014. ISSN 1432-1106. doi:10.1007/s00221-014-3895-z.
- [2] V. C. Paulun, K. R. Gegenfurtner, M. A. Goodale, and R. W. Fleming. Effects of material properties and object orientation on precision grip kinematics. *Experimental Brain Research*, 234(8):2253–2265, Aug. 2016. ISSN 1432-1106. doi:10.1007/s00221-016-4631-7.
- [3] R. W. Fleming, C. Wiebel, and K. Gegenfurtner. Perceptual qualities and material classes. *Journal of Vision*, 13(8):9, July 2013. ISSN 1534-7362. doi:10.1167/13.8.9.
- [4] R. W. Fleming. Visual perception of materials and their properties. *Vision Research*, 94:62–75, Jan. 2014. ISSN 0042-6989. doi:10.1016/j.visres.2013.11.004.
- [5] R. W. Fleming. Material Perception. *Annual Review of Vision Science*, 3(Volume 3, 2017):365–388, Sept. 2017. ISSN 2374-4642, 2374-4650. doi:10.1146/annurev-vision-102016-061429.
- [6] A. C. Zoeller, A. Lezkan, V. C. Paulun, R. W. Fleming, and K. Drewing. Integration of prior knowledge during haptic exploration depends on information type. *Journal of Vision*, 19(4): 20, Apr. 2019. ISSN 1534-7362. doi:10.1167/19.4.20.
- [7] G. Maiello, M. Schepko, L. K. Klein, V. C. Paulun, and R. W. Fleming. Humans Can Visually Judge Grasp Quality and Refine Their Judgments Through Visual and Haptic Feedback. *Frontiers in Neuroscience*, 14, Jan. 2021. ISSN 1662-453X. doi:10.3389/fnins.2020.591898.
- [8] M. Li, L. Zhang, T. Li, and Y. Jiang. Learning Gentle Grasping from Human-Free Force Control Demonstration. *IEEE Robotics and Automation Letters*, 10(3):2391–2398, Mar. 2025. ISSN 2377-3766, 2377-3774. doi:10.1109/LRA.2025.3530354.
- [9] K. Nakahara and R. Calandra. Learning Gentle Grasping Using Vision, Sound, and Touch. In *2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 7607–7614, Oct. 2025. doi:10.1109/IROS60139.2025.11246722.
- [10] J. Xu, M. Danielczuk, J. Ichnowski, J. Mahler, E. Steinbach, and K. Goldberg. Minimal Work: A Grasp Quality Metric for Deformable Hollow Objects. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1546–1552, May 2020. doi:10.1109/ICRA40945.2020.9197062.
- [11] I. Huang, Y. Narang, R. Bajcsy, F. Ramos, T. Hermans, and D. Fox. DefGraspNets: Grasp Planning on 3D Fields with Graph Neural Nets. <https://arxiv.org/abs/2303.16138v1>, Mar. 2023.
- [12] H. Zhang, K. Dana, and K. Nishino. Reflectance Hashing for Material Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3071–3080, 2015.
- [13] K. J. Dana, B. Van Ginneken, S. K. Nayar, and J. J. Koenderink. Reflectance and texture of real-world surfaces. *ACM Transactions on Graphics*, 18(1):1–34, Jan. 1999. ISSN 0730-0301, 1557-7368. doi:10.1145/300776.300778.

- [14] M. Weinmann, J. Gall, and R. Klein. Material Classification Based on Training Data Synthesized Using a BTF Database. In D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, editors, *Computer Vision – ECCV 2014*, volume 8691, pages 156–171. Springer International Publishing, Cham, 2014. ISBN 978-3-319-10577-2 978-3-319-10578-9. doi:10.1007/978-3-319-10578-9_11.
- [15] E. H. Adelson. On seeing stuff: The perception of materials by humans and machines. In *Human Vision and Electronic Imaging VI*, volume 4299, pages 1–12. SPIE, June 2001. doi:10.1117/12.429489.
- [16] L. Sharan, C. Liu, R. Rosenholtz, and E. H. Adelson. Recognizing Materials Using Perceptually Inspired Features. *International Journal of Computer Vision*, 103(3):348–371, July 2013. ISSN 1573-1405. doi:10.1007/s11263-013-0609-0.
- [17] S. Bell, P. Upchurch, N. Snavely, and K. Bala. Material Recognition in the Wild With the Materials in Context Database. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3479–3487, 2015.
- [18] K. J. Dana. *Computational Texture and Patterns: From Textons to Deep Learning*. Synthesis Lectures on Computer Vision. Springer International Publishing, Cham, 2018. ISBN 978-3-031-00695-1 978-3-031-01823-7. doi:10.1007/978-3-031-01823-7.
- [19] L. Liu, J. Chen, P. Fieguth, G. Zhao, R. Chellappa, and M. Pietikäinen. From BoW to CNN: Two Decades of Texture Representation for Texture Classification. *International Journal of Computer Vision*, 127(1):74–109, Jan. 2019. ISSN 1573-1405. doi:10.1007/s11263-018-1125-z.
- [20] G. Schwartz and K. Nishino. Recognizing Material Properties from Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(8):1981–1995, Aug. 2020. ISSN 1939-3539. doi:10.1109/TPAMI.2019.2907850.
- [21] M. Purri and K. Dana. Teaching Cameras to Feel: Estimating Tactile Physical Properties of Surfaces from Images. In A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, editors, *European Conference on Computer Vision*, volume 12372, pages 1–20. Springer International Publishing, Cham, 2020. ISBN 978-3-030-58582-2 978-3-030-58583-9. doi:10.1007/978-3-030-58583-9_1.
- [22] M. Beveridge and S. K. Nayar. Hierarchical Material Recognition from Local Appearance. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8165–8176, 2025.
- [23] R. L. Klatzky, D. K. Pai, and E. P. Krotkov. Perception of Material from Contact Sounds. *Presence: Teleoperators and Virtual Environments*, 9(4):399–410, Aug. 2000. ISSN 1054-7460. doi:10.1162/105474600566907.
- [24] C. Fang, D. Wang, F. Guo, J. Zou, and D. Song. A Fingertip Sensor and Algorithms for Pre-touch Distance Ranging and Material Detection in Robotic Grasping, Nov. 2023.
- [25] T. Zhang, M. Sheinin, D. Chan, M. Rau, M. O’Toole, and S. G. Narasimhan. Analyzing Physical Impacts Using Transient Surface Wave Imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4339–4348, 2023.
- [26] M. K. Johnson and E. H. Adelson. Retrographic sensing for the measurement of surface texture and shape. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1070–1077, June 2009. doi:10.1109/CVPR.2009.5206534.
- [27] M. K. Johnson, F. Cole, A. Raj, and E. H. Adelson. Microgeometry capture using an elastomeric sensor. *ACM Transactions on Graphics (TOG)*, 30(4):46:1–46:8, July 2011. ISSN 0730-0301. doi:10.1145/2010324.1964941.

- [28] W. Yuan, S. Dong, and E. H. Adelson. GelSight: High-Resolution Robot Tactile Sensors for Estimating Geometry and Force. *Sensors*, 17(12):2762, Dec. 2017. ISSN 1424-8220. doi:[10.3390/s17122762](https://doi.org/10.3390/s17122762).
- [29] M. Lambeta, P.-W. Chou, S. Tian, B. Yang, B. Maloon, V. R. Most, D. Stroud, R. Santos, A. Byagowi, G. Kammerer, D. Jayaraman, and R. Calandra. DIGIT: A Novel Design for a Low-Cost Compact High-Resolution Tactile Sensor With Application to In-Hand Manipulation. *IEEE Robotics and Automation Letters*, 5(3):3838–3845, July 2020. ISSN 2377-3766. doi:[10.1109/LRA.2020.2977257](https://doi.org/10.1109/LRA.2020.2977257).
- [30] W. Yuan, C. Zhu, A. Owens, M. A. Srinivasan, and E. H. Adelson. Shape-independent hardness estimation using deep learning and a GelSight tactile sensor. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 951–958, May 2017. doi:[10.1109/ICRA.2017.7989116](https://doi.org/10.1109/ICRA.2017.7989116).
- [31] W. Yuan, R. Li, M. A. Srinivasan, and E. H. Adelson. Measurement of shear and slip with a GelSight tactile sensor. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pages 304–311, Seattle, WA, USA, May 2015. IEEE. ISBN 978-1-4799-6923-4. doi:[10.1109/ICRA.2015.7139016](https://doi.org/10.1109/ICRA.2015.7139016).
- [32] R. Bhirangi, T. Hellebrekers, C. Majidi, and A. Gupta. ReSkin: Versatile, replaceable, lasting tactile skins. In *Proceedings of the 5th Conference on Robot Learning*, pages 587–597. PMLR, Jan. 2022.
- [33] B. Huang and Y. Li. FlexiTac: A Low-Cost, Open-Source, Scalable Tactile Sensing Solution for Robotic Systems, Apr. 2026.
- [34] V. Pattabiraman, Z. Huang, D. Panozzo, D. Zorin, L. Pinto, and R. Bhirangi. eFlesh: Highly customizable Magnetic Touch Sensing using Cut-Cell Microstructures, June 2025.
- [35] B. Huang, Y. Wang, X. Yang, Y. Luo, and Y. Li. 3D-ViTac: Learning Fine-Grained Manipulation with Visuo-Tactile Sensing. <https://arxiv.org/abs/2410.24091v2>, Oct. 2024.
- [36] X. Zhu, B. Huang, and Y. Li. Touch in the Wild: Learning Fine-Grained Manipulation with a Portable Visuo-Tactile Gripper. <https://arxiv.org/abs/2507.15062v2>, July 2025.
- [37] W. Yuan, S. Wang, S. Dong, and E. Adelson. Connecting Look and Feel: Associating the Visual and Tactile Properties of Physical Materials. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5580–5588, 2017.
- [38] Y. Li, J.-Y. Zhu, R. Tedrake, and A. Torralba. Connecting Touch and Vision via Cross-Modal Prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10609–10618, 2019.
- [39] X. Yu, R. Huang, C. Zhao, L. Zhou, and L. Ou. Def-Grasp: A Robot Grasping Detection Method for Deformable Objects Without Force Sensor. *Neural Processing Letters*, 55(8): 11739–11756, Dec. 2023. ISSN 1573-773X. doi:[10.1007/s11063-023-11398-8](https://doi.org/10.1007/s11063-023-11398-8).
- [40] W. Ruan, W. Zhu, K. Wang, Q. Lu, W. Yeh, L. Luo, C. Su, and Q. Wang. Vision-Tactile Fusion Based Detection of Deformation and Slippage of Deformable Objects During Grasping. In F. Sun, A. Cangelosi, J. Zhang, Y. Yu, H. Liu, and B. Fang, editors, *Cognitive Systems and Information Processing*, pages 593–604, Singapore, 2023. Springer Nature. ISBN 978-981-9906-17-8. doi:[10.1007/978-981-99-0617-8_43](https://doi.org/10.1007/978-981-99-0617-8_43).
- [41] N. Hogan. Impedance Control: An Approach to Manipulation: Part II—Implementation. *Journal of Dynamic Systems, Measurement, and Control*, 107(1):8–16, Mar. 1985. ISSN 0022-0434. doi:[10.1115/1.3140713](https://doi.org/10.1115/1.3140713).

- [42] T. Li, C. Chi, C. Wang, N. Xue, C. Liu, and C. Liu. Real-Time and Autonomous Grasping Operation of Manipulator Based on Tactile Sensor Array. In *2018 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*, pages 732–736, May 2018. doi:10.1109/IMCEC.2018.8469277.
- [43] S. J. Dharbaneshwer and A. Thondiyath. Contact Area-Based Modeling of Robotic Grasps Using Deformable Solid Mechanics. *International Journal of Applied Mechanics*, 13(03): 2150038, Apr. 2021. ISSN 1758-8251. doi:10.1142/S1758825121500381.
- [44] S. D’Avella, M. Fontana, R. Vertechy, and P. Tripicchio. Towards autonomous soft grasping of deformable objects using flexible thin-film electro-adhesive gripper. In *2022 IEEE 18th International Conference on Automation Science and Engineering (CASE)*, pages 1309–1314, Aug. 2022. doi:10.1109/CASE49997.2022.9926531.
- [45] J. W. James and N. F. Lepora. Slip Detection for Grasp Stabilization With a Multifingered Tactile Robot Hand. *IEEE Transactions on Robotics*, 37(2):506–519, Apr. 2021. ISSN 1941-0468. doi:10.1109/TRO.2020.3031245.
- [46] E. Judd, B. Aksoy, K. M. Digumarti, H. Shea, and D. Floreano. Slip Anticipation for Grasping Deformable Objects Using a Soft Force Sensor. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 10003–10008, Oct. 2022. doi:10.1109/IROS47612.2022.9981174.
- [47] T. N. Le, J. Lundell, and V. Kyrki. Safe Grasping with a Force Controlled Soft Robotic Hand. In *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 342–349, Oct. 2020. doi:10.1109/SMC42975.2020.9283496.
- [48] H. Dong, C.-Y. Chen, C. Qiu, C.-H. Yeow, and H. Yu. GSG: A Granary Soft Gripper with Mechanical Force Sensing via 3-Dimensional Snap-Through Structure, Nov. 2021.
- [49] T. N. Le, J. Lundell, F. J. Abu-Dakka, and V. Kyrki. Towards synthesizing grasps for 3D deformable objects with physics-based simulation. <https://arxiv.org/abs/2107.08898v1>, July 2021.
- [50] T. N. Le, J. Lundell, F. J. Abu-Dakka, and V. Kyrki. Deformation-Aware Data-Driven Grasp Synthesis. *IEEE Robotics and Automation Letters*, 7(2):3038–3045, Apr. 2022. ISSN 2377-3766. doi:10.1109/LRA.2022.3146551.
- [51] X. Lin, C. Qi, Y. Zhang, Z. Huang, K. Fragkiadaki, Y. Li, C. Gan, and D. Held. Planning with Spatial-Temporal Abstraction from Point Clouds for Deformable Object Manipulation. In *Proceedings of The 6th Conference on Robot Learning*, pages 1640–1651. PMLR, Mar. 2023.
- [52] K. Zhang, B. Li, K. Hauser, and Y. Li. AdaptiGraph: Material-Adaptive Graph-Based Neural Dynamics for Robotic Manipulation, July 2024.
- [53] K. Zhang, B. Li, K. Hauser, and Y. Li. Particle-Grid Neural Dynamics for Learning Deformable Object Models from RGB-D Videos, Nov. 2025.
- [54] H. Jiang, H.-Y. Hsu, K. Zhang, H.-N. Yu, S. Wang, and Y. Li. PhysTwin: Physics-Informed Reconstruction and Simulation of Deformable Objects from Videos. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7219–7230, 2025.
- [55] H. Zhang, G. Zheng, K. Zhang, J. Song, S. Patel, S. Hu, Y. Li, C. Zheng, and P. Y. Chen. BoxTwin: Learning Elastoplastic Articulated Object Dynamics from Videos.
- [56] R. Calandra, A. Owens, D. Jayaraman, J. Lin, W. Yuan, J. Malik, E. H. Adelson, and S. Levine. More Than a Feeling: Learning to Grasp and Regrasp using Vision and Touch. *IEEE Robotics and Automation Letters*, 3(4):3300–3307, Oct. 2018. ISSN 2377-3766, 2377-3774. doi:10.1109/LRA.2018.2852779.

- [57] S. Cui, R. Wang, J. Wei, F. Li, and S. Wang. Grasp State Assessment of Deformable Objects Using Visual-Tactile Fusion Perception. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 538–544, May 2020. doi:[10.1109/ICRA40945.2020.9196787](https://doi.org/10.1109/ICRA40945.2020.9196787).
- [58] Y. Luo, C. Liu, Y. J. Lee, J. DelPreto, K. Wu, M. Foshey, D. Rus, T. Palacios, Y. Li, A. Torralba, and W. Matusik. Adaptive tactile interaction transfer via digitally embroidered smart gloves. *Nature Communications*, 15(1):868, Jan. 2024. ISSN 2041-1723. doi:[10.1038/s41467-024-45059-8](https://doi.org/10.1038/s41467-024-45059-8).
- [59] Y. Wi, J. Yin, E. Xiang, A. Sharma, J. Malik, M. Mukadam, N. Fazeli, and T. Hellebrekers. TactAlign: Human-to-Robot Policy Transfer via Tactile Alignment. <https://arxiv.org/abs/2602.13579v1>, Feb. 2026.
- [60] E. Helmut, N. Funk, T. Schneider, C. de Farias, and J. Peters. Tactile-Conditioned Diffusion Policy for Force-Aware Robotic Manipulation. <https://arxiv.org/abs/2510.13324v1>, Oct. 2025.
- [61] B. Calli, A. Walsman, A. Singh, S. Srinivasa, P. Abbeel, and A. M. Dollar. Benchmarking in Manipulation Research: The YCB Object and Model Set and Benchmarking Protocols. <https://arxiv.org/abs/1502.03143v1>, Feb. 2015.
- [62] H. Liang, X. Ma, S. Li, M. Görner, S. Tang, B. Fang, F. Sun, and J. Zhang. PointNetGPD: Detecting Grasp Configurations from Point Sets. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 3629–3635, May 2019. doi:[10.1109/ICRA.2019.8794435](https://doi.org/10.1109/ICRA.2019.8794435).
- [63] A. Mousavian, C. Eppner, and D. Fox. 6-DOF GraspNet: Variational Grasp Generation for Object Manipulation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2901–2910, 2019.
- [64] X. Yan, J. Hsu, M. Khansari, Y. Bai, A. Pathak, A. Gupta, J. Davidson, and H. Lee. Learning 6-DOF Grasping Interaction via Deep Geometry-Aware 3D Representations. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3766–3773, May 2018. doi:[10.1109/ICRA.2018.8460609](https://doi.org/10.1109/ICRA.2018.8460609).
- [65] H.-S. Fang, C. Wang, M. Gou, and C. Lu. GraspNet-1Billion: A Large-Scale Benchmark for General Object Grasping. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11444–11453, 2020.
- [66] A. X. Chang, T. Funkhouser, L. Guibas, P. Hanrahan, Q. Huang, Z. Li, S. Savarese, M. Savva, S. Song, H. Su, J. Xiao, L. Yi, and F. Yu. ShapeNet: An Information-Rich 3D Model Repository. <https://arxiv.org/abs/1512.03012v1>, Dec. 2015.
- [67] L. Downs, A. Francis, N. Koenig, B. Kinman, R. Hickman, K. Reymann, T. B. McHugh, and V. Vanhoucke. Google Scanned Objects: A High-Quality Dataset of 3D Scanned Household Items. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 2553–2560, May 2022. doi:[10.1109/ICRA46639.2022.9811809](https://doi.org/10.1109/ICRA46639.2022.9811809).
- [68] M. Deitke, D. Schwenk, J. Salvador, L. Weihs, O. Michel, E. VanderBilt, L. Schmidt, K. Ehsani, A. Kembhavi, and A. Farhadi. Objaverse: A Universe of Annotated 3D Objects. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13142–13153, 2023.
- [69] J. Cao and E. Kalogerakis. SOPHY: Learning to Generate Simulation-Ready Objects with Physical Materials. <https://arxiv.org/abs/2504.12684v3>, Apr. 2025.
- [70] I. Huang, Y. Narang, C. Eppner, B. Sundaralingam, M. Macklin, R. Bajcsy, T. Hermans, and D. Fox. DefGraspSim: Physics-Based Simulation of Grasp Outcomes for 3D Deformable Objects. *IEEE Robotics and Automation Letters*, 7(3):6274–6281, July 2022. ISSN 2377-3766. doi:[10.1109/LRA.2022.3158725](https://doi.org/10.1109/LRA.2022.3158725).

- [71] F. P. Bowden and D. Tabor. *The Friction and Lubrication of Solids*. Clarendon Press, 2001. ISBN 978-0-19-850777-2.
- [72] D. A. H. Hanaor, Y. Gan, and I. Einav. Static friction at fractal interfaces. *Tribology International*, 93:229–238, Jan. 2016. ISSN 0301-679X. doi:[10.1016/j.triboint.2015.09.016](https://doi.org/10.1016/j.triboint.2015.09.016).
- [73] H. Hertz. On the contact of elastic solids. In *Miscellaneous Papers*, chapter 5, pages 146–162. Macmillan, London, 1896.
- [74] H. Hertz. On the contact of rigid elastic solids and on hardness. In *Miscellaneous Papers*, chapter 6, pages 163–183. Macmillan, London, 1896.
- [75] A. N. Gent. On the Relation between Indentation Hardness and Young’s Modulus. *Rubber Chemistry and Technology*, 31(4):896–906, Sept. 1958. ISSN 1943-4804, 0035-9475. doi:[10.5254/1.3542351](https://doi.org/10.5254/1.3542351).
- [76] A. W. Mix and A. J. Giacomini. Standardized Polymer Durometry. *Journal of Testing and Evaluation*, 39(4):696–705, July 2011. ISSN 0090-3973. doi:[10.1520/JTE103205](https://doi.org/10.1520/JTE103205).
- [77] M. Li, Z. Ferguson, T. Schneider, T. Langlois, D. Zorin, D. Panozzo, C. Jiang, and D. M. Kaufman. Incremental potential contact: Intersection-and inversion-free, large-deformation dynamics. *ACM Transactions on Graphics*, 39(4), Aug. 2020. ISSN 0730-0301, 1557-7368. doi:[10.1145/3386569.3392425](https://doi.org/10.1145/3386569.3392425).
- [78] K. Huang, F. Chitalu, H. Lin, and T. Komura. GIPC: Fast and stable Gauss-Newton optimization of IPC barrier energy. *ACM Transactions on Graphics*, 43(2):1–18, Apr. 2024. ISSN 0730-0301, 1557-7368. doi:[10.1145/3643028](https://doi.org/10.1145/3643028).
- [79] K. Huang, X. Lu, H. Lin, T. Komura, and M. Li. StiffGIPC: Advancing GPU IPC for Stiff Affine-Deformable Simulation. *ACM Transactions on Graphics*, 44(3):31:1–31:20, May 2025. ISSN 0730-0301. doi:[10.1145/3735126](https://doi.org/10.1145/3735126).
- [80] C. M. Kim, M. Danielczuk, I. Huang, and K. Goldberg. IPC-GraspSim: Reducing the Sim2Real Gap for Parallel-Jaw Grasping with the Incremental Potential Contact Model, Mar. 2022.
- [81] D. H. Nguyen, T. Schneider, G. Duret, A. Kshirsagar, B. Belousov, and J. Peters. TacEx: GelSight Tactile Simulation in Isaac Sim – Combining Soft-Body and Visuotactile Simulators, Nov. 2024.
- [82] R. Hasani, M. Lechner, A. Amini, L. Liebenwein, A. Ray, M. Tschaikowski, G. Teschl, and D. Rus. Closed-form continuous-time neural networks. *Nature Machine Intelligence*, 4(11): 992–1003, Nov. 2022. ISSN 2522-5839. doi:[10.1038/s42256-022-00556-7](https://doi.org/10.1038/s42256-022-00556-7).
- [83] C. Ferrari and J. Canny. Planning optimal grasps. In *Proceedings 1992 IEEE International Conference on Robotics and Automation*, pages 2290–2295 vol.3, May 1992. doi:[10.1109/ROBOT.1992.219918](https://doi.org/10.1109/ROBOT.1992.219918).
- [84] A. Miller and P. Allen. Graspit! A versatile simulator for robotic grasping. *IEEE Robotics & Automation Magazine*, 11(4):110–122, Dec. 2004. ISSN 1558-223X. doi:[10.1109/MRA.2004.1371616](https://doi.org/10.1109/MRA.2004.1371616).
- [85] J. M. Romano, K. Hsiao, G. Niemeyer, S. Chitta, and K. J. Kuchenbecker. Human-inspired robotic grasp control with tactile sensing. *IEEE Transactions on Robotics*, 27(6):1067–1079, 2011.
- [86] F. Veiga, H. Van Hoof, J. Peters, and T. Hermans. Stabilizing novel objects by learning to predict tactile slip. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5065–5072. IEEE, 2015.

- [87] S. K. Nayar, G. Krishnan, M. D. Grossberg, and R. Raskar. Fast separation of direct and global components of a scene using high frequency illumination. In *ACM SIGGRAPH 2006 Papers*, pages 935–944. 2006.

How Firm Should a Grasp Be?

Supplementary Material

S1 Methods Used to Collect the *Squash* Dataset

To capture the geometry and texture of each object, we used a Shining 3D Einscan SP V2 structured light scanner which has a single shot accuracy of ≤ 0.05 mm. The objects are first placed in an initial orientation on the turntable, and rotated to capture the full surface using between 8–36 views, depending on the texture of the object. The object is then reoriented at least one additional time, and the process is repeated to ensure comprehensive coverage. The resulting 3D meshes are postprocessed to be watertight, and have high resolution to capture the fine details in the object’s shape and texture. The 3D scan additionally provides a measurement of the object’s volume, and by weighing the object with a kitchen scale to get its mass, we can calculate its density.

To measure the object’s elastic modulus, we perform a uniaxial compression test with a parallel-jaw gripper. As force is applied to the object, the resulting deformation is observed by tactile sensors at the jaw fingertips. The relationship between the applied force and the resulting deformation is used to calculate the effective elastic modulus of both the tactile sensor (GelSight gel pad) and object. As shown in Fig. S1a, we have several measurements of the effective modulus during closure, which we average for a final estimate. The object’s elastic modulus is then recovered through calibration of the tactile fingertip’s own elastic modulus, *i.e.*, by using a rigid standard object. The object’s elastic modulus is reported as a function of an assumed Poisson ratio (Fig. S1a), since we can not directly measure it using our robotic setup.

To measure the object’s (static) surface friction μ and interfacial shear strength τ , we perform a tribological “drop test”. The object starts lifted and the applied force is reduced until the object slips, which is detected through optical flow in the tactile sensor. The applied force and contact area at the moment of slip are then used to calculate μ and τ through the soft friction model in Eq. (1). Several trials are performed per object, each at a different grasp location. As seen in Fig. S1b, the multiple samples allow us to fit a linear model to recover μ and τ during estimation.

S2 Deformable Grasp Simulation Details

Below we provide additional details on our deformable grasp simulation, including the physics engine, scene configuration, control strategy for data generation, and the data types that are captured during the course of a simulated grasp. We use this simulation to generate 200k training grasp sequences for our visuotactile model ψ , which estimates the material properties of the grasped object.

S2.1 Physics Engine and Constitutive Model

We simulate robotic grasps using UIPC [78, 79], a GPU-accelerated Incremental Potential Contact (IPC) solver [77], which has been modified to additionally model the effects of adhesion. In *Squash*, we have triangular surface meshes captured through structured light scanning. We first ensure each surface mesh is manifold and non-self-intersecting, then tetrahedralise the surface meshes prior to simulation. Each simulated grasped object is treated as a deformable body, and we model them with a Stable Neo-Hookean constitutive law and discretized as first-order tetrahedral finite elements (FEM). In our case, since our target real-world robot platform uses GelSight tactile sensors as the gripper fingertips, we also model the gripper fingertips (gel pads) as deformable bodies in the same manner. Contact is resolved with IPC’s barrier-based formulation, which guarantees intersection-free trajectories throughout the simulation. The contact distance threshold is set to $d=1$ mm, and a simulation time step of $\Delta t=4$ ms is used throughout.

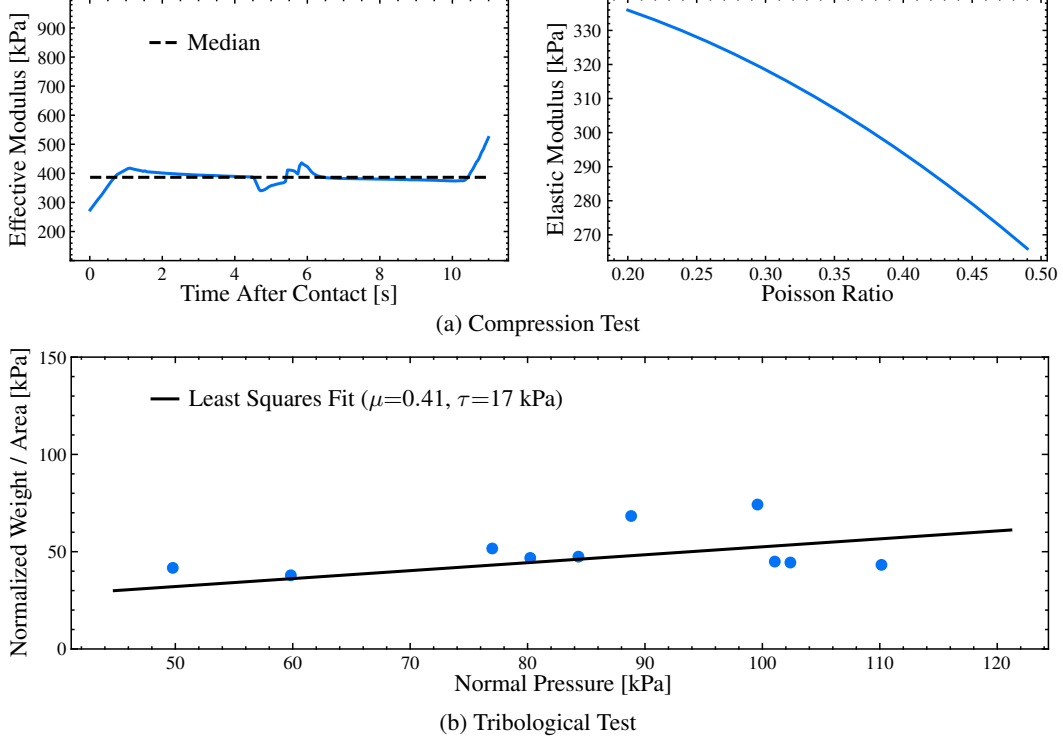


Figure S1. **Extracting Material Properties from Probing Tests.** The above plots show the probing results for a single instance of a peach. (a) A compression test is performed to measure the elastic modulus of the object, which is reported as a function of the Poisson ratio. (b) A tribological test is performed to measure the friction coefficient and shear strength of the object, where each point corresponds to a different trial at a different location on the object’s surface. The fitted line is used to decouple the two properties during estimation.

S2.2 Scene Configuration

Each simulation scene contains three bodies: a left gripper finger, a right gripper finger, and a grasped object. The gripper fingers are high-resolution meshes of the GelSight Mini gel pad. They are assigned an elastic modulus of $E_{\text{grip}}=245$ kPa, Poisson ratio $\nu_{\text{grip}}=0.48$, mass $m_{\text{grip}}=3$ g, friction coefficient $\mu_{\text{grip}}=0.9$, and interfacial shear strength $\tau_{\text{grip}}=100$ kPa, chosen to match the silicone elastomer of the physical sensor.

The grasped object is assigned material properties (E, m, μ, τ) that constitute the estimation target of the learned model. These properties are drawn from per-class Gaussian distributions fitted to the empirical property measurements in *Squash*. Poisson ratio ν is not directly measured in *Squash*; instead we fit a curve which models the relationship between elastic modulus E and an assumed value for ν . We thus assume all objects have $\nu=0.48$, and adjust our value for E accordingly. Inter-object friction and contact compliance are computed with the harmonic mean and the Hertz effective modulus, respectively, ensuring physical consistency at each contact pair.

To maximize coverage of contact geometry, each object is randomly oriented before placement. A random rotation is sampled uniformly from $[0, 2\pi)^3$ Euler angles, and the object is subsequently aligned so that a randomly chosen in-plane direction (lying in the yz -plane) coincides with the object’s principal geometric axis. In addition, the object shape is augmented to vary its dimensions. The object is then centered at the scene origin and the gel pads are positioned symmetrically about it along the x -axis, with an initial gap just beyond the contact distance threshold $\hat{d}$.

S2.3 Simulated Gripper Control

We employ an *adaptive force-controlled* gripper. During the *approach* phase, both jaws close at a constant speed of $v_{\text{close}}=4$ mm s⁻¹ until the contact force on either jaw exceeds a contact detection threshold of 0.25 N, or until the jaw displacement reaches 30% of the object’s half-width plus $\hat{d}$

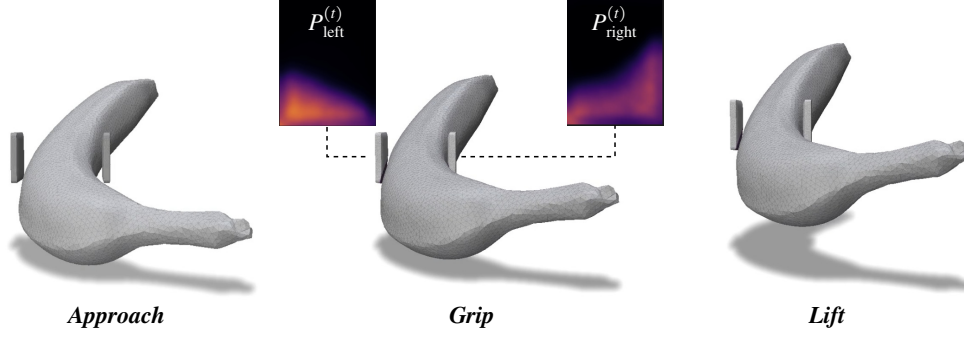


Figure S2. **Simulating Deformable Object Grasps.** We use UIPC, a GPU-accelerated incremental potential contact solver, to simulate the dynamics of a parallel-jaw gripper with deformable elastomeric gel pads grasping a deformable object. Deformable objects are taken from the *Squash* dataset, and their material properties and dimensions are augmented to vary the contact geometry. The simulator extracts rich visuotactile observations at each time step, such as the pressure map $P^{(t)}$ observed in the local frame of the jaw pads, which are used to train our learned model for material estimation and force control.

(whichever comes first). Once contact is detected (*i.e.*, the *grip* phase has begun), the jaws enter a proportional (P) control loop that adjusts their displacement to drive the measured contact force F_{load} toward a target grip force F^* . The control law is given by:

$$x \leftarrow x + K_p (F^* - F_{\text{load}}) \text{sgn}(v_{\text{close}}), \quad K_p = 10^{-4} \text{ m N}^{-1}. \quad (\text{S1})$$

Transition to the *lift* phase is triggered once both jaws have been simultaneously in servo for at least five time-steps *and* the force error is within 20% of F^* , or after a maximum of 100 servo steps. After a brief 100 ms dwell, the gripper lifts at $v_{\text{lift}} = 10 \text{ mm s}^{-1}$ over a distance of 2 mm. This process is outlined in Fig. S2.

S2.4 Sensor Observation

At every time step, the simulator extracts the following tactile observations from each gripper jaw:

- **Scalar contact force** $F_{\text{load}} \in \mathbb{R}_{\geq 0} [\text{N}]$: the L^2 norm of the summed contact gradient vectors over all contact vertices, divided by Δt^2 .
- **Force vector** $\mathbf{F}_{\text{load}} \in \mathbb{R}^3 [\text{N}]$: the vector sum of per-vertex contact forces in the gripper jaw's local coordinate frame.
- **Total contact area** $a [\text{m}^2]$: the total area of surface triangles on the gripper jaw pad that contain at least one vertex in contact with the object.
- **Contact centroid** $\mathbf{c} \in \mathbb{R}^3$: the mean position of contact vertices, normalized by the bounding box of the gripper jaw pad and expressed relative to the pad center.
- **Gripper displacement** $\mathbf{d} \in \mathbb{R}^3 [\text{m}]$: the rigid-body displacement of the pad from its initial rest position at the beginning of the simulated grasp.
- **Force map (tactile image)** $I \in \mathbb{R}^{H \times W} [\text{N}]$: an image in which each pixel accumulates the contact force contribution of overlapping surface triangles, rasterized in the gripper jaw pad's yz -plane.
- **Pressure map** $P \in \mathbb{R}^{H \times W} [\text{Pa}]$: per-vertex contact pressure (area-weighted average of incident triangle pressures), rendered with barycentric interpolation into the same image space as the force map. An example of the pressure map is shown in Fig. S2.
- **Indentation map** $L \in \mathbb{R}^{H \times W} [\text{m}]$: per-vertex surface indentation depth, computed as the positive part of the difference between rest and deformed z -coordinates in the pad's body frame, and rendered with barycentric interpolation.
- **Contact area map** $A \in \{0, 1\}^{H \times W}$: binary mask indicating pixels with non-zero force.
- **Shear map** $V \in \mathbb{R}^{H \times W} [\text{m}]$: per-vertex shear displacement, computed as the L^2 norm of the yz -components of the vertex displacement, and rendered with barycentric interpolation.

- **Slip map** $U \in \mathbb{R}^{H \times W}$ [m]: per-vertex slip displacement, computed as the difference between the current and previous time step’s contact centroids, and rendered with barycentric interpolation.

The complete time-series of these observations, together with the object’s mechanical properties and grip force used in that trial, is stored as a compressed NumPy archive (.npz).

S2.5 Parallel and Distributed Execution

To amortize GPU setup overhead, we batch multiple grasps into a single UIPC scene by tiling simulations on a 2-D grid in the xz -plane. A data manifest specifies all combinations of object mesh, material properties, and grip force to be simulated; this manifest is partitioned into contiguous chunks and each chunk is dispatched to a separate GPU subprocess, enabling large-scale parallel data collection across multi-GPU compute nodes.

S2.6 IPC Simulation Sensitivity

To evaluate our simulator’s fidelity for grasping, we ran a sensitivity study over the IPC solver parameters. While per-object target grasp forces shift modestly (~ 0.1 N), our method’s MAE moves by $\leq 5\%$. Force estimates are insensitive to loosening friction regularization ($\epsilon_v \times 10$: -5%), halving/doubling the contact barrier (-4%), and halving the time step dt ($+3\%$). Larger ϵ_v or dt tend to reduce μ and τ identifiability, as incipient slip becomes less apparent.

S3 Visuotactile Model Design and Training Details

We now discuss the architecture of our learned model ψ for estimating material properties from visuotactile sequences, as well as details of its training and deployment. The model consists of a sequence encoder that processes sequential tactile observations and control signals, a likelihood network that scores candidate material parameter vectors, and an online particle filter that maintains a belief distribution over the latent material parameters.

S3.1 Model Architecture

Feature Representation. We represent the material properties of each object as a physical parameter vector $\theta = (E, m, \mu, \tau)$, where E is elastic modulus, m is mass, μ is the Coulomb friction coefficient, and τ is the interfacial shear strength. For inference, we work in a log-transformed latent space:

$$\mathbf{z} = [\ln E, \ln m, \ln \mu, \ln \tau] \in \mathbb{R}^4, \quad (\text{S2})$$

which maps the physically constrained parameters onto an unconstrained real-valued vector suitable for particle belief filtering. The inverse mapping $\mathbf{z} \rightarrow \theta$ applies elementwise exponentials.

At each time step t the robot’s tactile sensors (one per jaw) yield a tactile image $I^{(t)} \in \mathbb{R}^{H \times W}$. Each jaw’s image is encoded by a shared convolutional backbone: three strided Conv2d layers (channels $1 \rightarrow 16 \rightarrow 32 \rightarrow 64$, kernel sizes $5/3/3$, stride 2) followed by adaptive average pooling and a linear projection to an embedding of dimension $d_{\text{emb}}=128$. The two jaw embeddings are concatenated and fused by a single linear layer to produce a per-timestep feature vector $\mathbf{x}_{\text{feat}}^{(t)} \in \mathbb{R}^{d_{\text{emb}}}$. We then assemble a 28-dimensional control vector comprising raw signals (per-jaw normal force, jaw displacement in 3D, contact area, mean contact pressure, and contact centroid in 3D; 18 dimensions total) and their first-order temporal differences (force, area, and centroid; 10 dimensions). These are projected to $\mathbb{R}^{d_{\text{emb}}}$ by a two-layer MLP to yield $\mathbf{x}_{\text{ctrl}}^{(t)}$.

Sequence Encoder. The concatenated feature and control embedding $\mathbf{x}^{(t)} = [\mathbf{x}_{\text{feat}}^{(t)}; \mathbf{x}_{\text{ctrl}}^{(t)}] \in \mathbb{R}^{2d_{\text{emb}}}$ is processed by a Closed-form Continuous-time (CfC) recurrent network [82]. Concretely, given the sequence $(\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(t)})$, the CfC produces hidden states:

$$\mathbf{h}^{(t)} = \text{CfC}(\mathbf{x}^{(t)}, \mathbf{h}^{(t-1)}), \quad \mathbf{h}^{(t)} \in \mathbb{R}^{d_h}, \quad (\text{S3})$$

with hidden dimension $d_h=256$. The network is strictly causal: $\mathbf{h}^{(t)}$ depends only on past and present observations, enabling online inference.

Likelihood Network. The hidden state $\mathbf{h}^{(t)}$ is used to score candidate parameter vectors via a learned likelihood head. A candidate $\tilde{\mathbf{z}} \in \mathbb{R}^4$ is first projected to $\mathbb{R}^{d_{\text{emb}}}$ by a two-layer MLP. The projected candidate and the hidden state are concatenated and passed through a two-layer scoring head MLP that outputs an unnormalised log-likelihood score $s = \log p_\psi(I^{(0:t)}, F_{\text{load}}^{(0:t)} | \tilde{\mathbf{z}})$. A learnable log-temperature parameter (initialized to $\log 10$) scales the logits.

Particle Filter. The particle filter maintains a set of $N=512$ weighted particles $\{(\theta_j^{(t)}, \alpha_j^{(t)})\}_{j=1}^N$ representing the posterior over $\mathbf{z}$ (equivalently, over θ). Particles are initialized by sampling from the class prior $p_c(\theta)$. At each time step, the filter executes the following steps:

1. **Diffuse.** A random-walk perturbation is applied to each particle to model process noise:

$$\theta_j^{(t)} \leftarrow \theta_j^{(t-1)} + \epsilon_j, \quad \epsilon_j \sim \mathcal{N}(\mathbf{0}, \sigma_{\text{rw}}^2 \mathbf{I}), \quad (\text{S4})$$

and particles are clipped to physically valid latent-space bounds.

2. **Score.** The likelihood network evaluates $s_j = \log p_\psi(I^{(t)}, F_{\text{load}}^{(t)} | \theta_j^{(t)})$ for all N particles in a single batched forward pass through $\mathbf{h}^{(t)}$.

3. **Reweight.** Log-sum-exp normalisation yields importance weights:

$$\alpha_j^{(t)} \propto \alpha_j^{(t-1)} \exp(s_j). \quad (\text{S5})$$

4. **Resample.** If the effective sample size $\text{ESS} = (\sum_j (\alpha_j^{(t)})^2)^{-1}$ falls below $N/2$, systematic resampling is performed and weights are reset to $1/N$.

The weighted mean $\hat{\mathbf{z}}^{(t)} = \sum_j \alpha_j^{(t)} \mathbf{z}_j^{(t)}$ is decoded to physical space to yield the point estimate $\hat{\theta}^{(t)}$, which is used to compute $F_{\text{min}}^{(t)}$ at each time step (the minimum required per-jaw normal force) by solving the Hertz-coupled friction balance (see Eqs. (1) and (2) in the main paper).

S3.2 Model Training from Precomputed Simulations

The network is trained on 200k simulated grasping sequences with ground-truth material parameters, as described in Sec. 5 of the main paper and Sec. S2 above. Each sequence consists of T timesteps of tactile images and applied loads paired with a fixed ground-truth $\mathbf{z}_{\text{gt}}$.

We use a hybrid contrastive-regression objective. The primary term is an InfoNCE loss: at each time step t , the hidden state $\mathbf{h}^{(t)}$ scores one positive ($\mathbf{z}_{\text{gt}}$) against $K=127$ negatives sampled from the per-object prior, plus all other in-batch ground-truth vectors as cross-negatives. The loss for batch element b at timestep t is:

$$\mathcal{L}_{b,t}^{\text{NCE}} = -\log \frac{\exp(\beta \cdot s^+)}{\exp(\beta \cdot s^+) + \sum_k \exp(\beta \cdot s_k^-)}, \quad (\text{S6})$$

where β is a learned temperature scale. A secondary MSE regression term directly predicts $\mathbf{z}_{\text{gt}}$ from $\mathbf{h}^{(t)}$, weighted by $\lambda=0.1$. Both terms are averaged using a binary mask $\mathbf{v}_{b,t} \in \{0, 1\}$:

$$\mathcal{L} = \frac{\sum_{b,t} \mathbf{v}_{b,t} (\mathcal{L}_{b,t}^{\text{NCE}} + \lambda \mathcal{L}_{b,t}^{\text{reg}})}{\sum_{b,t} \mathbf{v}_{b,t}}. \quad (\text{S7})$$

The network is trained with Adam ($\text{lr}=10^{-4}$) for 500 epochs with a batch size of 8 sequences using a cosine learning rate schedule. Training is distributed across 8 NVIDIA A6000 Ada GPUs, taking approximately 48 hours to complete.

S4 Additional Experimental Results

Below, we provide additional evaluations of our approach when grasping simulated and real objects.

Table S1. **Efficacy of Material-Aware Robot Grasping in Simulation (Continuation of Tab. 1).** We compare our material-aware approach for gripper force control with two standard approaches using 5k simulated produce items with high variability. “Naïve” uses a force of 3 N in each gripper jaw, which just exceeds the threshold to successfully lift the largest object in *Squash*. “Class Prior” uses the class’s average material properties to determine the grasp force using Eq. (1). In contrast, our approach refines instance-level mechanical properties to yield improved fine-grained force control. We additionally report the performance of our approach when removing the class prior (from *observation*), and when the class prior is incorrectly identified. We see that even without an accurate estimate of the object’s class-level material prior, our approach achieves gentle grasps relative to the non-adaptive baselines. The best results are in bold, and second best are underlined.

Grasp Force Controller	Material estimation MAPE ↓				Force error ↓		
	E	m	μ	τ	Overload [N]	Work [mJ]	Impulse [Ns]
<i>Open-Loop Baseline Grasping Approaches (Non-Adaptive)</i>							
Naïve constant force (3 N)	—	—	—	—	2.01 ± 0.41	3.32 ± 2.04	2.15 ± 1.28
Class prior	1.07 ± 0.25	1.50 ± 2.13	1.21 ± 0.41	0.77 ± 0.32	1.63 ± 1.42	3.63 ± 5.76	1.95 ± 2.58
<i>Material-Aware Grasping Approaches (Adaptive), with Ablations</i>							
Ours	0.56 ± 0.23	0.49 ± 0.24	0.56 ± 0.25	0.44 ± 0.17	0.39 ± 0.26	2.25 ± 5.37	0.66 ± 0.90
↳ no prior	0.65 ± 0.81	0.62 ± 0.97	0.60 ± 0.39	0.58 ± 0.20	0.44 ± 0.27	2.65 ± 1.97	0.73 ± 0.62
↳ wrong (adversarial) prior	0.69 ± 0.45	<u>0.56</u> ± 1.04	0.63 ± 0.31	<u>0.55</u> ± 0.19	<u>0.42</u> ± 0.33	2.99 ± 6.76	1.44 ± 1.12

Error Metrics. We measure the performance of our visuotactile model using several metrics. Bias measures, in Newtons, whether the approach tends to undershoot or overshoot. Mean absolute error (MAE) measures the absolute error in Newtons. Weighted absolute percentage error (WAPE) and mean absolute percentage error (MAPE) measure performance on a normalized scale. Overload reports the peak force exerted above the optimal force, and is in essence a measure of overshoot. Excess work and excess impulse convey the risk of damage to the object. Success rate is the percentage of grasps where the object did not fall from the grasp of the gripper after it has been lifted.

Results and Ablations in Simulation. In our simulation experiments, our visuotactile model achieved a high success rate in predicting optimal grasp forces for a wide range of deformable objects. This is seen in Fig. S3, where, on average, our approach exerts a much lower force than baseline methods to successfully complete a grasp. The visuotactile model was able to adapt its predictions based on the material properties of the objects, such as their elasticity, mass, and friction coefficients, as shown in Tab. S1. The results showed that our model outperformed baseline methods that did not utilize tactile feedback, demonstrating the importance of incorporating tactile information for gentle grasping of deformable objects. In 83.0% of cases, our approach yielded the most optimal grasp force, compared to 15.4% for the *Class Prior* approach and 1.6% for the *Naïve* approach. In addition to our stated method of refining class-level material priors to instance-level posteriors, we perform two ablation tests. In the first, we removed the class-level material prior entirely. In the second, we set the class-prior to an incorrect distribution for the object being grasped (*e.g.*, using the prior for apples when grasping an apricot). We observed that our approach maintains the ability to reasonably estimate material properties from tactile feedback, and thus achieve gentle grasps in both ablation tests.

Results in Real-World Experiments. In our real-world experiments, our visuotactile model demonstrated a significant improvement in grasping performance compared to baseline methods. The model was able to successfully grasp a variety of test objects, applying appropriate forces that minimized the risk of damage to the objects. In Fig. S4, we show the distribution of estimation errors over the range of material properties seen in the real-world test set. While our approach performs well on average, as one would expect, we see degraded performance when the material properties are significantly different from the class prior. This matches what we observed in our ablation studies above, where we removed or adversarially adjusted the class prior.

S5 Robot Hardware and Operating System

For both data collection and real-world validation of material-aware grasping, we utilize a 6 DoF Universal Robots UR5 arm. The arm is equipped with a Weiss Robotics WSG50-110 parallel-jaw gripper, which has a maximum jaw opening of 110 mm and spatial resolution of 0.1 mm. Each jaw is fitted with a 2 kg bending-beam load cell to measure the axial force applied by each gripper jaw

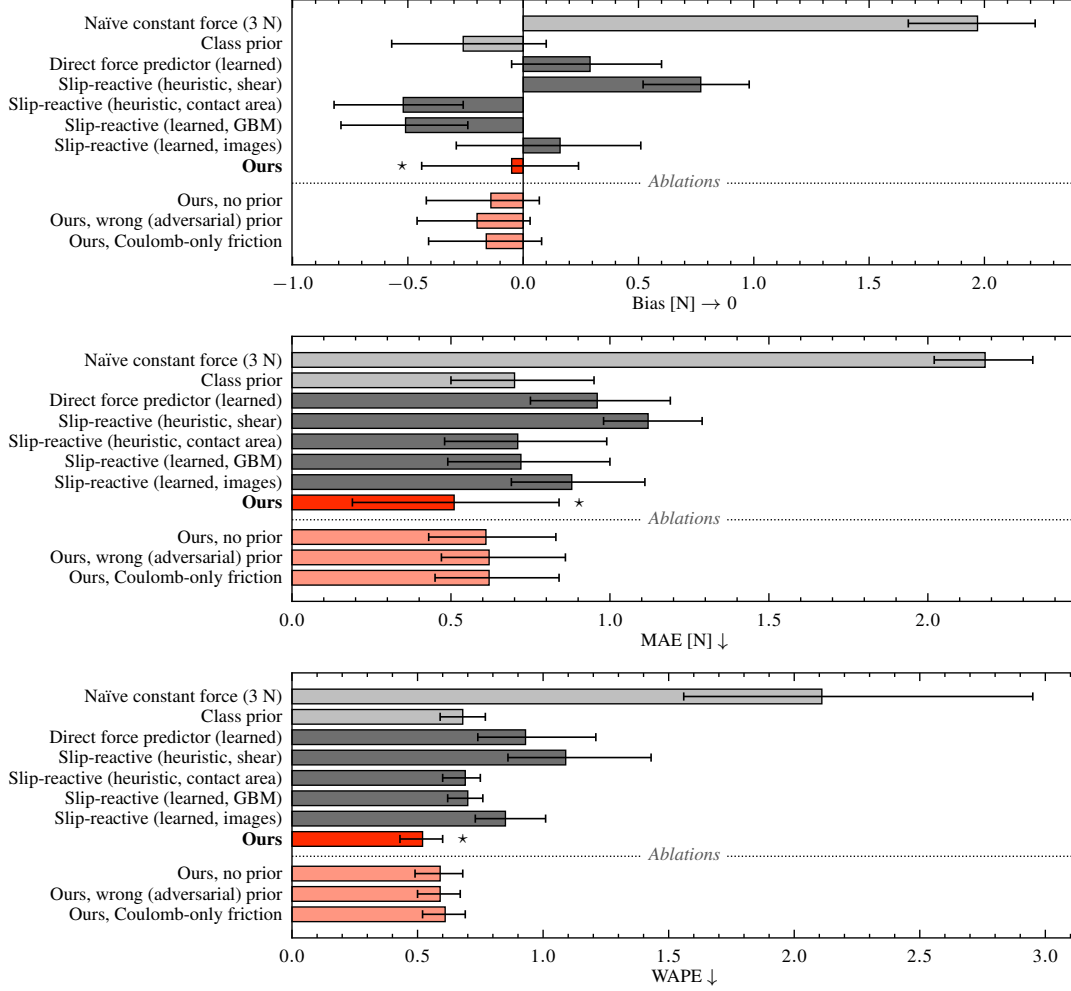


Figure S3. **Force Estimation Error on Simulated Object Grasps.** The bar plots show the (top) signed error/bias, (middle) mean absolute error, and (bottom) weighted absolute percentage error of the estimated minimum required grip force $F_{\min}$ compared to the ground truth, across all simulated grasping sequences. Error bars show the confidence interval. The error is computed once the object has been lifted by 2 mm, which is 0.2 s after lifting begins. The best approach is indicated by a *. The results demonstrate that our visuotactile model is able to accurately estimate the required grip force for a wide range of deformable objects with varying material properties.

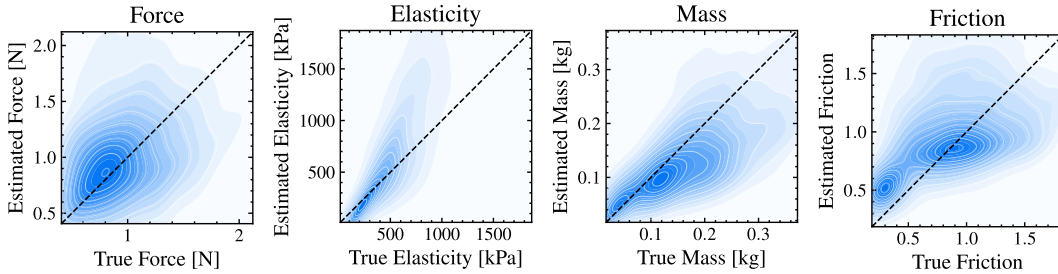


Figure S4. **Heatmap of Force and Material Estimation Errors for Real Object Grasps.** The estimated force and material properties are compared against the ground truth measurements for a set of real-world grasps of fruits and vegetables. The heatmap visualizes the distribution of estimation errors of our approach. Note that the heatmap color is indicative of the range of material properties observed in the test set, and the extent of the boundary represents the range of estimation errors using a kernel density estimate. In general, we see our approach achieves low estimation errors but may struggle at the extremes of the material property range, far from the average of the class prior.

to the object being grasped. The load cells utilize a Wheatstone bridge powered by a clean external power supply. The input voltage is regulated by LT3042 voltage regulators to ensure consistent performance, and the output voltage is amplified by INA333 voltage amplifiers. We read the differential voltage between input and output at 50 kHz using MCC128 DAQ hats on a Raspberry Pi 5b. At the end of each load cell, we mount a GelSight Mini tactile sensor, which provides freerunning high-resolution 8 MP tactile images at a rate of 25 Hz.

For real-world validation of our material-aware grasping approach, we implement a ROS node that runs our learned visuotactile model for force control at 250 Hz. The node subscribes to the tactile image and load cell data streams, processes the data to extract the necessary features, and outputs the target grip force to the gripper control loop. We use a simple PID controller to servo the gripper towards the target force. Since GelSight Mini operates at a lower frame rate, we interpolate the tactile images to match the control frequency. The ROS node is implemented in Python and C++, and utilizes PyTorch for running the learned model inference. A desktop computer is used to read the tactile sensor data and run the control node, and it is PTP synchronized with the Raspberry Pi 5b that records and publishes the load cell measurements. We ensure that the entire pipeline from sensor data acquisition to control output operates within the required real-time constraints for effective grasping.