

Dex-Net 3.0: Computing Robust Robot Suction Grasp Targets in Point Clouds using a New Analytic Model and Deep Learning

Jeffrey Mahler¹, Matthew Matl¹, Xinyu Liu¹, Albert Li¹, David Gealy¹, Ken Goldberg^{1,2}

Abstract—Suction-based end effectors are widely used in industry and are often preferred over parallel-jaw and multifinger grippers due to their ability to lift objects with a single point of contact. This ability simplifies planning, and hand-coded heuristics such as targeting planar surfaces are often used to select suction grasps based on point cloud data. In this paper, we propose a compliant suction contact model that computes the quality of the seal between the suction cup and target object and determines whether or not the suction grasp can resist an external wrench (e.g. gravity) on the object. To evaluate a grasp, we measure robustness to perturbations in end-effector and object pose, material properties, and external wrenches. We use this model to generate Dex-Net 3.0, a dataset of 2.8 million point clouds, suction grasps, and grasp robustness labels computed with 1,500 3D object models and we train a Grasp Quality Convolutional Neural Network (GQ-CNN) on this dataset to classify suction grasp robustness from point clouds. We evaluate the resulting system in 375 physical trials on an ABB YuMi fitted with a pneumatic suction gripper. When the object shape, pose, and mass properties are known, the model achieves 99% precision on a dataset of objects with Adversarial geometry such as sharply curved surfaces. Furthermore, a GQ-CNN-based policy trained on Dex-Net 3.0 achieves 99% and 97% precision respectively on a dataset of Basic and Typical objects. Code, datasets, and supplemental material can be found at <http://berkeleyautomation.github.io/dex-net>.

I. INTRODUCTION

Suction grasping is a widely-used method for pick-and-place tasks in industry and warehouse order fulfillment. As shown in the Amazon Picking Challenge, suction has an advantage over parallel-jaw or multifinger grasping due to its ability to reach into narrow spaces and pick up objects with a single point of contact. However, while a substantial body of research exists on parallel-jaw and multifinger grasp planning [3], comparatively little research exists on planning suction grasps.

While grasp planning searches for gripper configurations that maximize a quality metric derived from mechanical wrench space analysis [28], human labels [32], or self-supervised labels [22], suction grasps are often planned directly on point clouds using heuristics such as grasping near the object centroid [12] or targeting planar surfaces [6], [4]. These heuristics work well for prismatic objects such as boxes and cylinders but may fail on objects with complex geometry. Some analytic models of suction cups exist, but they typically assume that a vacuum seal has already been formed and that the state (e.g. shape and pose) of the object

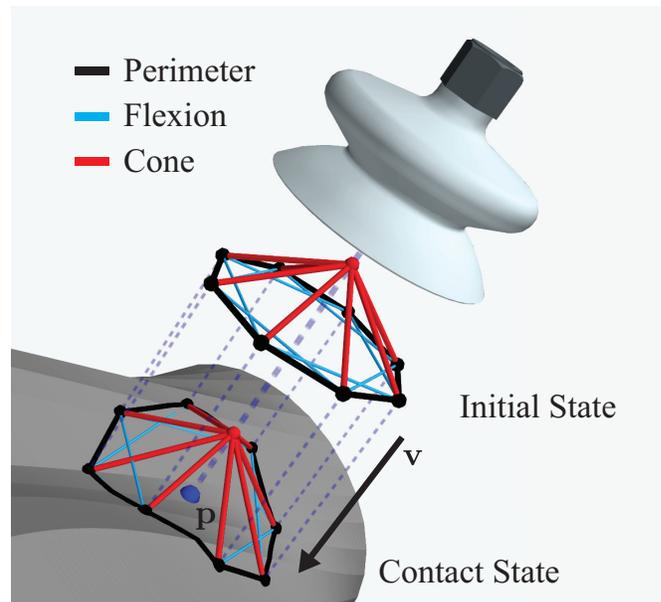


Fig. 1: The quasi-static spring model, C , used for determining when seal formation is feasible. The model contains three types of springs – perimeter, flexion, and cone springs. An initial state for C is chosen given a target point p and an approach direction v . Then, a contact state for C is computed so that C 's perimeter springs form a complete seal against object mesh M . Seal formation is deemed feasible if the energy required to maintain this contact state is sufficiently low in every spring.

is perfectly known [2], [19], [26]. In practice a robot may need to form seals on non-planar surfaces while being robust to external wrenches (e.g. gravity and disturbances), sensor noise, control imprecision, and calibration errors, which are significant factors when planning grasps from point clouds.

We propose a novel compliant suction contact model for rigid, non-porous objects that consists of two components: (1) a test for whether a seal can be formed between a suction cup and a target object and (2) an analysis of the ability of the suction contact to resist external wrenches. We use the model to evaluate grasp robustness by evaluating seal formation and wrench resistance under perturbations in object pose, suction tip pose, material properties, and disturbing wrenches using Monte-Carlo sampling similar to that in the Dexterity Network (Dex-Net) 1.0 [25].

This paper makes five contributions:

- 1) A compliant suction contact model that measures seal formation using a quasi-static spring system and measures the ability to resist external wrenches (e.g. gravity) using a contact wrench basis derived from the ring of contact between the cup and object surface.
- 2) *Robust wrench resistance*: a robust version of the above model under random disturbing wrenches and

¹ Dept. of Electrical Engineering and Computer Science;

² Dept. of Industrial Operations and Engineering Research; AUTOLAB University of California, Berkeley, USA {jmahler, mmatl, xinyuliu, alberhli, dgealy, goldberg}@berkeley.edu

- perturbations in object pose, gripper pose, and friction.
- 3) Dex-Net 3.0, a dataset of 2.8 million synthetic point clouds annotated with suction grasps and grasp robustness labels generated by analyzing robust wrench resistance for approximately 375k grasps across 1,500 object models.
 - 4) Experiments characterizing the precision of robust wrench resistance both with and without knowledge of the target object’s shape and pose.
 - 5) A taxonomy of object shapes for robot grasping experiments that differentiates 3 classes of objects (Basic, Typical, and Adversarial).

We perform physical experiments using an ABB YuMi robot with a silicone suction cup tip to compare the precision of a GQ-CNN-based grasping policy trained on Dex-Net 3.0 against several heuristics such as targeting planar surfaces, and we find that our method achieves 99% and 97% average precision on the Basic and Typical objects, respectively.

II. RELATED WORK

End-effectors based on suction are widely used in industrial applications such as warehouse order fulfillment, handling limp materials such as fabric [19] or produce [27], and robot applications such as the Amazon Picking Challenge [4], underwater manipulation [34], or wall climbing [2]. Our method builds on models of deformable materials, analyses of the wrenches that suction cups can exert, and data-driven grasp planning.

A. Suction Models

Several models for the deformation of stiff rubber-like materials exist in the literature. Provot et al. [30] modeled sheets of stiff cloth and rubber using a spring-mass system with several types of springs. Hosseini et al. [1] provided a survey of more modern constitutive models of rubber that are often used in Finite Element Analysis (FEA) packages for more realistic physics simulations. In order to rapidly evaluate whether a suction cup can form a seal against an object’s surface, we model the cup as a quasi-static spring system with a topology similar to the one in [30] and estimate the deformation energy required to maintain a seal.

In addition, several models have been developed to check for static equilibrium assuming a seal between the suction cup and the object’s surface. Most models consider the suction cup to be a rigid object and model forces into the object along the surface normal, tangential forces due to surface friction, and pulling forces due to suction [19], [34], [36], [40]. Bahr et al. [2] augmented this model with the ability of to resist moments about the center of suction to determine the amount of vacuum pressure necessary to keep a climbing robot attached to a vertical wall. Mantriota [26] modeled torsional friction due to a contact area between the cup and object similar to the soft finger contact model used in grasping [16]. Xin et al. [39] experimentally characterized the relationship between the suction pressure distribution, air power consumption, and gasket height, finding that suction force is approximately proportional to air power. Our model

extends these methods by combining models of torsional friction [26] and contact moments [2] in a compliant ring contact model.

B. Grasp Planning

The goal of grasp planning is to select a configuration for an end-effector that enables a robot to perform a task via contact with an object while resisting external perturbations [3], [33], which can be arbitrary [8] or task-specific [20], [23]. A common approach is to select a configuration that maximizes a quality metric (or reward) based on wrench space analysis [10], [28], robustness to perturbations [37], or a model learned from human labels [32] or self-supervision [29].

Several similar metrics exist for evaluating suction grasps. One common approach is to evaluate whether or not a set of suction cups can lift an object by applying an upwards force [19], [34], [35], [36]. Domae et al. [6] developed a geometric model to evaluate suction success by convolving target locations in images with a desired suction contact template to assess planarity. In addition, heuristics for planning suction grasps from point clouds have been used extensively in the Amazon Picking Challenge. In 2015, Team RBO [7] won by pushing objects from the top or side until suction was achieved using a vacuum with high air flow, and Team MIT [41] came in second place using a strategy that positioned the suction cup directly above the centroid of objects with flat surfaces. In 2016, Team Delft [12] won the challenge using a suction planning strategy that selected the closest point on the object surface to the object centroid and approached the point along the inward surface normal. In this work, we present a novel metric that evaluates whether a suction cup can resist external wrenches given a target point on a 3D model, and we estimate grasp robustness via Monte-Carlo sampling of wrench resistance over perturbations in poses, material properties, and disturbing wrenches.

This paper also extends empirical, data-driven approaches to grasp planning based on images and point clouds [3]. A popular approach is to use human labels of graspable regions in RGB-D images [21] or point clouds [5], [17] to learn a grasp detector with computer vision techniques. As labeling may be tedious for humans, an alternative is to automatically collect training data from a physical robot and to train Convolutional Neural Network (CNN) to predict grasp success [22], [29]. To reduce the time-cost of data collection, recent research has proposed to generate labels in simulation using physical models of contact [14], [17]. Mahler et al. [24] demonstrated that a GQ-CNN trained on Dex-Net 2.0, a dataset of 6.7 million point clouds, grasps, and quality labels computed with robust quasi-static analysis, could be used to successfully plan grasps across a wide variety of objects with 99% precision. In this paper, we use a similar approach to generate a dataset of point clouds, grasps, and robustness labels for a suction-based end-effector.

III. PROBLEM STATEMENT

Given a point cloud from a depth camera, our goal is to find a robust suction grasp (target point and approach direction) for a robot to lift and transport a singulated object from a planar worksurface to a receptacle. We solve for the suction grasp that maximizes the probability that the robot can hold the object under perturbations due to sensor noise and disturbing wrenches using a stochastic model of grasp success and image formation.

A. Assumptions

Our stochastic model makes the following assumptions:

- 1) Quasi-static physics (e.g. inertial terms are negligible) with Coulomb friction.
- 2) Objects are rigid and made of non-porous material.
- 3) Each object is singulated on a planar worksurface in a stable resting pose [9].
- 4) A single overhead depth sensor oriented orthogonally to the worksurface with known intrinsics, position, and orientation relative to the robot.
- 5) A vacuum-based end-effector with known geometry and a single disc-shaped suction cup made of linear-elastic material.

B. Definitions

A robot observes a single-view **point cloud** or depth image, \mathbf{y} , containing a singulated object. The goal is to find the most robust suction **grasp** \mathbf{u} that enables the robot to lift an object and transport it to a receptacle, where grasps are parametrized by a target point $\mathbf{p} \in \mathbb{R}^3$ and an approach direction $\mathbf{v} \in \mathcal{S}^2$. Success is measured with a binary **grasp quality function** S , where $S = 1$ if the grasp \mathbf{u} successfully transports the object, and $S = 0$ otherwise.

The robot may not be able to predict the success of suction grasps exactly from point clouds for several reasons. First, the success metric depends on a latent **state** \mathbf{x} describing the object’s geometric, inertial, and material properties \mathcal{O} and the pose of the object relative to the camera, T_o , but the robot does not know the true state due to: (a) noise in the depth image and (b) occlusions due to the single viewpoint. Second, the robot may not have perfect knowledge of external wrenches (forces and torques) on the object due to gravity or external disturbances.

This probabilistic relationship is described by an **environment** consisting of a grasp success distribution modeling $p(S | \mathbf{u}, \mathbf{x})$, the ability of a grasp to resist random disturbing wrenches, and an observation model $p(\mathbf{y} | \mathbf{x})$. This model induces a probability of success for each grasp conditioned on the robot’s observation:

Definition 1: The *robustness* of a grasp \mathbf{u} given a point cloud \mathbf{y} with respect to an environment p is the probability of grasp success under uncertainty in sensing, control, and disturbing wrenches: $Q(\mathbf{u}, \mathbf{y}) = \mathbb{P}(S | \mathbf{u}, \mathbf{y})$.

Our environment model is described in Section V and further details are given in Appendix V.

C. Objective

Our goal is to find a grasp that maximizes robustness given a point cloud, $\pi^*(\mathbf{y}) = \operatorname{argmax}_{\mathbf{u} \in \mathcal{C}} Q(\mathbf{u}, \mathbf{y})$, where \mathcal{C} specifies constraints on the set of available grasps, such as collisions or kinematic feasibility. However, in practice the robot does not have an explicit representation of the robustness function Q but has the ability to sample from the stochastic environment. As inferring the state over a large number of possible objects may be challenging, we approximate π^* by training a deep Grasp Quality Convolutional Neural Network (GQ-CNN) $Q_\theta(\mathbf{u}, \mathbf{y})$ on samples of success labels, point clouds, and suction grasps from our model by minimizing the cross-entropy loss \mathcal{L} :

$$\min_{\theta \in \Theta} \sum_{i=1}^N \mathcal{L}(S_i, Q_\theta(\mathbf{u}_i, \mathbf{y}_i)) \quad (\text{III.1})$$

Learning Q rather than directly learning π^* allows us to enforce task-specific constraints without having to update the learned model. Given optimized GQ-CNN weights $\hat{\theta}$, we execute the deep robust grasping policy $\pi_\theta(\mathbf{y}) = \operatorname{argmax}_{\mathbf{u} \in \mathcal{C}} Q_{\hat{\theta}}(\mathbf{u}, \mathbf{y})$.

IV. COMPLIANT SUCTION CONTACT MODEL

To evaluate grasp success, we develop a novel quasi-static spring model of the suction cup material and a model of contact wrenches that can be applied through a ring of contact between the object and suction cup perimeter. Our success metric evaluates whether:

- 1) A seal is formed between the perimeter of the suction cup and the object surface.
- 2) Given a seal, the suction cup is able to resist an external wrench on the object due to gravity and disturbances.

A. Seal Formation

A suction cup lifts objects with an inwards force due to an air pressure differential induced across the membrane of the cup by a vacuum generator. If a gap exists between the perimeter of the cup and the object, then air flowing into the gap may reduce the differential and cause the grasp to fail. Therefore, forming a tight seal between the cup and the target object is crucial for grasp success.

To determine when seal formation is possible, we model the suction cup as a spring system C parameterized by real numbers (n, r, h) . Rather than performing a computationally expensive dynamic simulation with a spring-mass model to determine when seal formation is feasible, we make simplifying assumptions to evaluate seal formation geometrically. Specifically, we compute a configuration of C that achieves a seal by projecting C onto the surface of M and evaluate the feasibility of that configuration under quasi-static conditions as a proxy for the dynamic feasibility of seal formation.

In our model, C has two types of springs – *structural* springs that represent the physical structure of the suction cup and *flexion* springs that do not correspond to physical structures but instead are used to resist bending along the cup’s surface. Dynamic spring-mass systems with similar

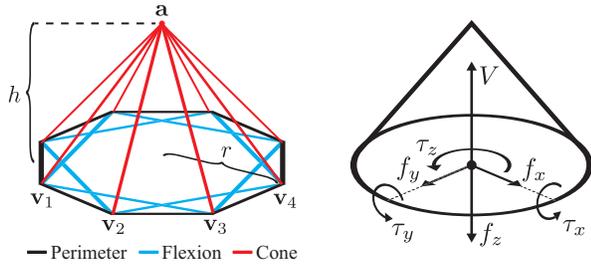


Fig. 2: Our compliant suction contact model. **(Left)** The quasi-static spring model used in seal formation computations. This suction cup is approximated by $n = 8$ components. Here, r is equal to the radius of the cup and h is equal to the height of the cup. $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ are the base vertices and \mathbf{a} is the apex. **(Right)** Wrench basis for the suction ring contact model. The contact exerts a constant pulling force on the object of magnitude V and additionally can push or pull the object along the contact z axis with force f_z . The suction cup material exerts a normal force $f_N = f_z + V$ on the object through a linear pressure distribution on the ring. This pressure distribution induces a friction limit surface bounding the set of possible frictional forces in the tangent plane $f_t = (f_x, f_y)$ and the torsional moment τ_z , and also induces torques τ_x and τ_y about the contact x and y axes due to elastic restoring forces in the suction cup material.

structures have been used in prior work to model stiff sheets of rubber [30]. The undeformed structural springs of C form a right pyramid with height h and with a base that is a regular n -gon with circumradius r (see Fig. 11). Let $\mathcal{V} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n, \mathbf{a}\}$ be the set of vertices of the undeformed right pyramid, where each \mathbf{v}_i is a base vertex and \mathbf{a} is the pyramid’s apex. We define the model’s set of springs as follows:

- **Perimeter (Structural) Springs** – Springs linking vertex \mathbf{v}_i to vertex \mathbf{v}_{i+1} , $\forall i \in \{1, \dots, n\}$.
- **Cone (Structural) Springs** – Springs linking vertex \mathbf{v}_i to vertex \mathbf{a} , $\forall i \in \{1, \dots, n\}$.
- **Flexion Springs** – Springs linking vertex \mathbf{v}_i to vertex \mathbf{v}_{i+2} , $\forall i \in \{1, \dots, n\}$.

In this model, a complete seal is formed between C and a target object parameterized by 3D mesh M if and only if each of the perimeter springs of C lies entirely on the surface of M . Given a target mesh M with a target grasp $\mathbf{u} = (\mathbf{p}, \mathbf{v})$ for the gripper, we choose an initial configuration of C such that C is undeformed and the approach line (\mathbf{p}, \mathbf{v}) passes through \mathbf{a} and is orthogonal to the base of C . Then, we make the following assumptions to determine a final static contact configuration of C that forms a complete seal against M (see Fig. 1):

- **The perimeter springs** of C must not deviate from the original undeformed regular n -gon when projected onto a plane orthogonal to \mathbf{v} . This means that their locations can be computed by projecting them along \mathbf{v} from their original locations onto the surface of M .
- **The apex, \mathbf{a} ,** of C must lie on the approach line (\mathbf{p}, \mathbf{v}) and, given the locations of C ’s base vertices, must also lie at a location that keeps the average distance along \mathbf{v} between \mathbf{a} and the perimeter vertices equal to h .

See Appendix II for additional details.

Given this configuration, we check its feasibility to determine if a complete seal is actually achievable. The configuration is feasible if:

- The cone faces of C cannot collide with M during approach or in the contact configuration.
- The surface of M has no holes within the contact ring traced out by C ’s perimeter springs.
- The energy required in each spring to maintain the contact configuration of C is below a threshold.

We threshold the energy in individual springs rather than the total energy for C because air gaps are usually caused by local geometry – for example, a small groove in the mesh’s surface would require a significant amount of deformation energy from small sections of the cup, but the overall deformation energy could still be low if the rest of the mesh is planar.

B. Wrench Space Analysis

To determine whether or not the suction cup can resist external wrenches such as gravity through contact, we analyze the set of wrenches that the suction cup can apply.

1) **Wrench Resistance:** The object wrench set for a grasp using a contact model with m basis wrenches is $\Lambda = \{\mathbf{w} \in \mathbb{R}^6 \mid \mathbf{w} = G\alpha \text{ for some } \alpha \in \mathcal{F}\}$, where $G \in \mathbb{R}^{6 \times m}$ is a set of m basis wrenches in the object coordinate frame, and $\mathcal{F} \subseteq \mathbb{R}^m$ is a set of constraints on contact wrench magnitudes [28].

Definition 2: A grasp \mathbf{u} achieves *wrench resistance* with respect to \mathbf{w} if $-\mathbf{w} \in \Lambda$ [20], [28].

We encode wrench resistance as a binary variable W such that $W = 1$ if \mathbf{u} resists \mathbf{w} and $W = 0$ otherwise.

2) **Suction Contact Model:** Many suction contact models acknowledge normal forces, vacuum forces, tangential friction, and torsional friction [2], [19], [26], [34] similar to a point contact with friction or soft finger model [28]. However, under this model, a single suction cup cannot resist torques about axes in the contact tangent plane, implying that any torque about such axes would cause the suction cup to drop an object (see Appendix III for a detailed proof). This defies our intuition since empirical evidence suggests that a single point of suction can robustly transport objects [7], [12].

We hypothesize that these torques are resisted through an asymmetric pressure distribution on the ring of contact between the suction cup and object, which occurs due to passive elastic restoring forces in the material. Fig. 11 illustrates the suction ring contact model. The grasp map G is defined by the following basis wrenches:

- 1) **Actuated Normal Force (f_z):** The force that the suction cup material applies to the object along the z axis.
- 2) **Vacuum Force (V):** The magnitude of the constant force pulling the object into the suction cup coming from the air pressure differential.
- 3) **Frictional Force ($f_f = (f_x, f_y)$):** The force in the contact tangent plane due to the normal force between the suction cup and object, $f_N = f_z + V$.
- 4) **Torsional Friction (τ_z):** The torque resulting from frictional forces in the ring of contact.
- 5) **Elastic Restoring Torque ($\tau_e = (\tau_x, \tau_y)$):** The torque about axes in the contact tangent plane resulting from

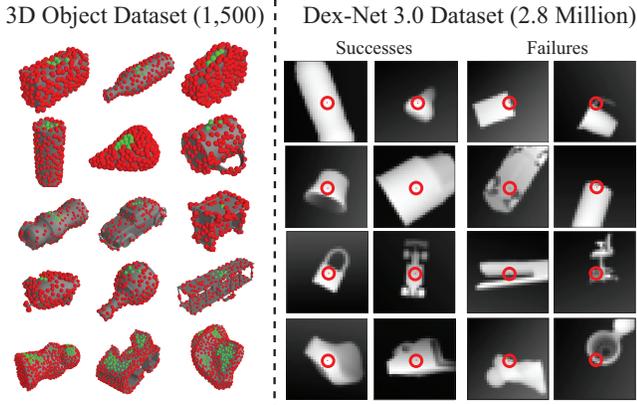


Fig. 3: The Dex-Net 3.0 dataset. (Left) The Dex-Net 3.0 object dataset contains approx. 350k unique suction target points across the surfaces of 1,500 3D models from the KIT object database [18] and 3DNet [38]. Each suction grasp is classified as robust (green) or non-robust (red). Robust grasps are often above the object center-of-mass on flat surfaces of the object. (Right) The Dex-Net 3.0 point cloud dataset contains approx. 2.8 million tuples of point clouds and suction grasps with robustness labels. The dataset contains approximately 11.8% positive examples.

elastic restoring forces in the suction cup pushing on the object along the boundary of the contact ring.

The magnitudes of the contact wrenches are constrained due to (a) the friction limit surface [16], (b) limits on the elastic behavior of the suction cup material, and (c) limits on the vacuum force. In the suction ring contact model, \mathcal{F} is defined by a set of linear constraints for efficient computation of wrench resistance:

$$\begin{aligned}
 \text{Friction:} \quad & \sqrt{3}|f_x| \leq \mu f_N \quad \sqrt{3}|f_y| \leq \mu f_N \quad \sqrt{3}|\tau_z| \leq r\mu f_N \\
 \text{Material:} \quad & \sqrt{2}|\tau_x| \leq \pi r \kappa \quad \sqrt{2}|\tau_y| \leq \pi r \kappa \\
 \text{Suction:} \quad & f_z \geq -V
 \end{aligned}$$

Here μ is the friction coefficient, r is the radius of the contact ring, and κ is a material-dependent constant. These constraints define a subset of the friction limit ellipsoid and cone of admissible elastic torques under a linear pressure distribution about the ring of the cup. Furthermore, we can compute wrench resistance using quadratic programming due to the linearity of the constraints. See Appendix III for a detailed derivation and proof.

C. Robust Wrench Resistance

We evaluate the robustness of candidate suction grasps by evaluating seal formation and wrench resistance over distributions on object pose, grasp pose, and disturbing wrenches:

Definition 3: The *robust wrench resistance* metric for \mathbf{u} and \mathbf{x} is $\rho(\mathbf{u}, \mathbf{x}) = q(W | \mathbf{u}, \mathbf{x})$ where q is a distribution over object poses, grasp poses, and disturbing wrenches.

In practice, we evaluate robust wrench resistance by taking M samples, evaluating wrench resistance for each, and computing the sample mean: $\sum_{j=1}^M W_i$.

V. DEX-NET 3.0 DATASET

To learn to predict grasp robustness based on noisy point clouds, we generate the Dex-Net 3.0 training dataset of point clouds, grasps, and grasp success labels by sampling tuples $(S_i, \mathbf{u}_i, \mathbf{y}_i)$ from a joint distribution $p(S, \mathbf{u}, \mathbf{x}, \mathbf{y})$ that is composed of distributions on:

- **States:** $p(\mathbf{x})$: A prior on possible objects, object poses, and camera poses that the robot will encounter.
- **Grasp Candidates:** $p(\mathbf{u}|\mathbf{x})$: A prior constraining grasp candidates to target points on the object surface.
- **Grasp Successes** $p(S | \mathbf{u}, \mathbf{x})$: A stochastic model of wrench resistance for the gravity wrench.
- **Observations** $p(\mathbf{y} | \mathbf{x})$: A sensor noise model.

To sample from the model, we first sample a latent state by selecting an object at random from a database of 3D CAD models and sampling a friction coefficient, planar pose of the object, and camera height above the table. We generate a set of grasp candidates for the object model by sampling points and normals uniformly at random from the surface of the object mesh. We then evaluate robust wrench resistance (ρ) as described in Section IV-C and convert to a binary label $S = \mathbf{1}(\rho > \epsilon)$ for a threshold $\epsilon = 0.2$. Finally, we sample a point cloud of the scene using rendering and a model of image noise [25]. The grasp success labels are associated with pixel locations in images through perspective projection [11]. Fig. 3 illustrates a subset of the Dex-Net 3.0 object and grasp dataset. Our pipeline for computing grasp robustness and generating training datapoints is illustrated in Fig. 4. For more details on the distributions, see Appendix V.

VI. LEARNING A DEEP ROBUST GRASPING POLICY

We use the Dex-Net 3.0 dataset to train a GQ-CNN that takes as input a single-view point cloud of an object resting on the table and a candidate suction grasp defined by a target 3D point and approach direction, and outputs the robustness, or estimated probability of success for the grasp on the visible object.

Our GQ-CNN architecture is identical to Dex-Net 2.0 [24] except that we modify the pose input stream to include the angle between the approach direction and the table normal and change the pc1 layer from size 16 to 64. The point cloud stream takes a depth image centered on the target point and rotated to align the middle column of pixels with the approach orientation similar to a spatial transforming layer [13]. The end-effector depth from the camera and orientation are input to a fully connected layer in a separate pose stream and concatenated with conv features in a fully connected layer. We train the GQ-CNN with using stochastic gradient descent with momentum using an 80-20 training-to-validation split of the Dex-Net 3.0 dataset. Training took approximately 12 hours on three NVIDIA Titan X GPUs. The learned GQ-CNN achieves 93.5% classification accuracy on the held-out validation set. For more details, see Appendix IV

We use the GQ-CNN in a deep robust grasping policy to plan suction target grasps from point clouds on a physical robot. The policy uses the Cross Entropy Method

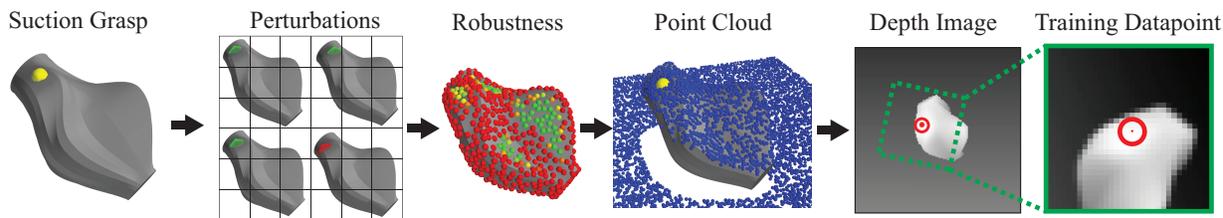


Fig. 4: Pipeline for generating the Dex-Net 3.0 dataset (left to right). We first sample a candidate suction grasp from the object surface and evaluate the ability to form a seal and resist gravity over perturbations in object pose, gripper pose, and friction. The samples are used to estimate the probability of success, or robustness, for candidate grasps on the object surface. We render a point cloud for each object and associate the candidate grasp with a pixel and orientation in the depth image through perspective projection. Training datapoints are centered on the suction target pixel and rotated to align with the approach axis to encode the invariance of the robustness to image locations.

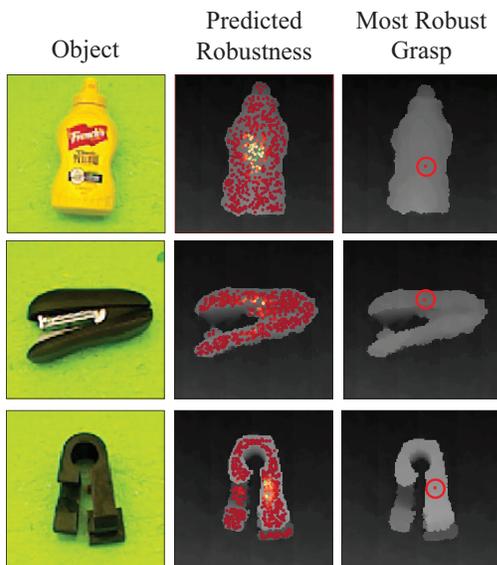


Fig. 5: Robust grasps planned with the Dex-Net 3.0 GQ-CNN-based policy on example RGB-D point clouds. **(Left)** The robot is presented an object in isolation. **(Middle)** Initial candidate suction target points colored by the predicted probability of success from zero (red) to one (green). Robust grasps tend to concentrate around the object centroid. **(Right)** The policy optimizes for the grasp with the highest probability of success using the Cross Entropy Method.

(CEM) [22], [24], [31]. CEM samples a set of initial candidate grasps uniformly at random from the set of surface points and inward-facing normals on a point cloud of the object, then iteratively resamples grasps from a Gaussian Mixture Model fit to the grasps with the highest predicted probability of success. Fig. 5 illustrates the probability of success predicted by the GQ-CNN on candidates grasps from several real point clouds.

VII. EXPERIMENTS

We ran experiments to characterize the precision of robust wrench resistance when object shape and pose are known and the precision of our deep robust grasping policy for planning grasps from point clouds for three object classes. For additional experiments, see Appendix I.

A. Object Classes

We created a dataset of 55 rigid and non-porous objects including tools, groceries, office supplies, toys, and 3D printed

industrial parts. We separated objects into three categories, illustrated in Fig. 6:

- 1) *Basic*: Prismatic solids (e.g. rectangular prisms, cylinders). Includes 25 objects.
- 2) *Typical*: Common objects with varied geometry and many accessible, approximately planar surfaces. Includes 25 objects.
- 3) *Adversarial*: 3D-printed objects with complex geometry (e.g. curved or narrow surfaces) that are difficult to access. Includes 5 objects.

For object details, see <http://bit.ly/2xMcx3x>.

B. Experimental Protocol

We ran experiments with an ABB YuMi with a Primesense Carmine 1.09 and a suction system with a 15mm silicone single-bellow suction cup and a VM5-NC vacuum generator with a payload of approximately 0.9kg. The experimental workspace is illustrated in the left panel of Fig. 6. In each experiment, the operator iteratively presented a target object to the robot and the robot planned and executed a suction grasp on the object. On each trial, the operator labeled successes based on whether or not the robot was able to lift and transport the object to the side of the workspace. For each method, we measured:

- 1) **Average Precision (AP)**. The area under the precision-recall curve, which measures precision over possible thresholds on the probability of success predicted by the policy. This is useful for industrial applications where a robot may take an alternative action (e.g. asking for help) if the planned grasp is predicted to fail.
- 2) **Success Rate**. The fraction of all grasps that were successful.

All experiments ran on a Desktop running Ubuntu 14.04 with a 2.7 GHz Intel Core i5-6400 Quad-Core CPU and an NVIDIA GeForce 980 GPU.

C. Performance on Known Objects

To assess performance of our robustness metric independently of the perception system, we evaluated whether or not the metric was predictive of suction grasp success when object shape and pose were known using the 3D printed Adversarial objects. The robot was presented one of the five



Fig. 6: (Left) The experimental setup with an ABB YuMi equipped with a suction gripper. (Right) The 55 objects used to evaluate suction grasping performance. The objects are divided into three categories to characterize performance: Basic (e.g. prismatic objects), Typical, and Adversarial.

Adversarial objects in a known stable pose, selected from the top three most probable stable poses. We hand-aligned the object to a template image generated by rendering the object in a known pose on the table. Then, we indexed a database of grasps precomputed on 3D models of the objects and executed the grasp with the highest metric value for five trials. In total, there were 75 trials per experiment.

We compared the following metrics:

- 1) **Planarity-Centroid (PC3D)**. The inverse distance to the object centroid for sufficiently planar patches on the 3D object surface.
- 2) **Spring Stretch (SS)**. The maximum stretch among virtual springs in the suction contact model.
- 3) **Wrench Resistance (WR)**.
- 4) **Robust Wrench Resistance (RWR)**.

The RWR metric performed best with 99% AP compared to 93% AP for WR, 89% AP for SS, and 88% for PC3D. Appendix I-B contains additional analyses.

D. Performance on Novel Objects

We also evaluated the performance of GQ-CNNs trained on Dex-Net 3.0 for planning suction target points from a single-view point cloud. In each experiment, the robot was presented one object from either the Basic, Typical, or Adversarial classes in a pose randomized by shaking the object in a box and placing it on the table. The object was imaged with a depth sensor and segmented using 3D bounds on the workspace. Then, the grasping policy executed the most robust grasp according to a success metric. In this experiment the human operators were blinded from the method they were evaluating to remove bias in human labels.

We compared policies that optimized the following metrics:

- 1) **Planarity**. The inverse sum of squared errors from an approach plane for points within a disc with radius equal to that of the suction cup.
- 2) **Centroid**. The inverse distance to the object centroid.
- 3) **Planarity-Centroid**. The inverse distance to the centroid for planar patches on the 3D object surface.
- 4) **GQ-CNN (ADV)**. A GQ-CNN trained on synthetic data from the Adversarial objects (to assess the ability of the model to fit complex objects).

- 5) **GQ-CNN (DN3)**. A GQ-CNN trained on synthetic data from 3DNet [38], KIT [18], and the Adversarial objects.

Table I details performance on the Basic, Typical, and Adversarial objects. We see that the Dex-Net 3.0 policy has the highest AP across the Basic and Typical classes. Also, the GQ-CNN trained on the Adversarial objects significantly outperforms all methods on the Adversarial dataset, suggesting that our model is able to exploit knowledge of complex 3D geometry to plan robust grasps. The Dex-Net 3.0 policy took an average of approximately 3 seconds to plan each grasp. Appendix I-C contains additional analyses.

E. Failure Modes

The mode common failure mode was attempting to form a seal on surfaces with small variations such as ridges and indentations that prevent a seal formation. We believe this is due to the limited resolution of the depth sensor, as our seal formation model is able to detect the inability to form a seal on such surfaces. The GQ-CNN-based policy also tends to fail on objects with small protrusions such as wall hooks. Appendix I-E contains additional information.

VIII. DISCUSSION AND FUTURE WORK

We presented a novel compliant model of suction contact and generated the Dex-Net 3.0 dataset from the model to train a deep CNN to predict grasp success from point clouds. Our experiments suggest that our suction contact model has high precision in predicting suction grasp success in physical trials.

In future work we plan to study the effect of the prior object dataset on system performance. We are interested in synthetic object dataset augmentation using different scales, mass distributions, and shapes. Furthermore, we plan to evaluate performance with higher resolution depth sensors, vacuums with higher payload, and suction cups with stiffer material and larger diameters. Finally, we are also interested in using the suction grasping policy as a primitive for bin picking, where a high-level policy decides between using parallel jaws or suction based on the predicted probability of success.

	Basic		Typical		Adversarial	
	AP (%)	Success Rate (%)	AP (%)	Success Rate (%)	AP (%)	Success Rate (%)
Planarity	81	74	69	67	48	47
Centroid	89	92	80	78	47	38
Planarity-Centroid	98	94	94	86	64	62
GQ-CNN (ADV)	83	77	75	67	86	81
GQ-CNN (DN3)	99	98	97	82	61	58

TABLE I: Performance of image-based grasping policies for 125 trials each on the Basic and Typical datasets and 100 trials each on the Adversarial datasets. We see that the GQ-CNN trained on Dex-Net 3.0 has the highest average precision on the Basic and Typical objects but has lower precision on the adversarial objects, which are very different than common objects in the training dataset. A GQ-CNN trained on the Adversarial dataset significantly outperforms all methods on these objects, suggesting that our model is able to capture complex geometries when the training dataset contains a large proportion of such objects.

ACKNOWLEDGMENTS

This research was performed at the AUTOLAB at UC Berkeley in affiliation with the Berkeley AI Research (BAIR) Lab, the Real-Time Intelligent Secure Execution (RISE) Lab, and the CITRIS People and Robots (CPAR) Initiative. The authors were supported in part by the U.S. National Science Foundation under NRI Award IIS-1227536: Multilateral Manipulation by Human-Robot Collaborative Systems, the Department of Defense (DoD) through the National Defense Science & Engineering Graduate Fellowship (NDSEG) Program, the Berkeley Deep Drive (BDD) Program, and by donations from Siemens, Knapp, Google, Cisco, Autodesk, IBM, Amazon Robotics, and Toyota Robotics Institute. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Sponsors. We thank our colleagues who provided helpful feedback, code, and suggestions, in particular Ruzena Bajcsy, Oliver Brock, Peter Corke, Chris Correa, Ron Fearing, Roy Fox, Menglong Guo, Michael Laskey, Andrew Lee, Pusong Li, Jacky Liang, Sanjay Krishnan, Fritz Kuttler, Stephen McKinley, Juan Aparicio Ojeda, Michael Peinhopf, Peter Puchwein, Alberto Rodriguez, Daniel Seita, Vishal Satish, and Shankar Sastry.

APPENDIX I

ADDITIONAL EXPERIMENTS

To better characterize the correlation of our robust wrench resistance metric, compliant suction contact model, and GQ-CNN-based policy for planning suction target grasps from point clouds with physical outcomes on a real robot, we present several additional analyses and experiments.

A. Success-Attempt Rate Tradeoff

Our primary numeric metrics of performance were:

- 1) **Average Precision (AP).** The area under the precision-recall curve, which measures precision over possible thresholds on the probability of success predicted by the policy. This is useful for industrial applications where a robot may take an alternative action (e.g. probing, asking for help) if the planned grasp is predicted to fail.
- 2) **Success Rate.** The fraction of all grasps that were successful.

We argue that these metrics alone do not give a complete picture of how useful a suction grasp policy would work in practice. Average Precision (AP) penalizes a policy for having poor recall (a high rate of false negatives relative to true positives), and success rate penalizes a policy with a high number of failures. However, not all failures should be treated equally: some failures are predicted to occur by the GQ-CNN (low predicted probability of success) while the others are the result of an overconfident prediction.

In practice, a suction grasp policy would be part of a larger system (e.g. a state machine) that could decide whether or not to execute a grasp based on the continuous probability of success output by the GQ-CNN. As long as the policy is not overconfident, such as system can detect failures before they occur and take an alternative action such as attempting to turn the object over, asking a human for help, or leaving the object in the bin for error handling. At the same time, if a policy is too conservative and never predicts successes, then the system will handle be able to handle very few test cases.

We illustrate this tradeoff by plotting the Success-Attempt Rate curve which plots:

- 1) **Success Rate.** The fraction of fraction of grasps that are successful if the system only executes grasps have predicted probability of success greater than a confidence threshold τ .
- 2) **Attempt Rate.** The fraction of all test cases for which the system attempts a grasp, if the system only attempts grasps with predicted probability of success greater than a confidence threshold τ .

over all possible values of the confidence threshold τ . There is typically an inverse relationship between the two metrics: a higher confidence threshold will reduce false positives but will also reduce the frequency of grasp attempts, increasing runtime and decreasing the diversity of cases that the robot is able to successfully handle.

B. Performance on Known Objects

See Section VII-C for a description of the experiment. Additional results are detailed in Table II and the Success-Attempt Rate curve is plotted in Fig. 7. A policy based on the robust wrench resistance metric achieved nearly 100% average precision and 92% success on this dataset, suggesting that the ranking of grasps by robust wrench resistance is correlated with the ranking by physical successes.

C. Performance on Novel Objects

See Section VII-D for a description of the experiment. Fig. 8 illustrates the Success-Attempt Rate tradeoff on the Basic, Typical, and Adversarial objects. We see that the Dex-Net 3.0 policy has the highest AP across the Basic

Metric	AP (%)	Success Rate (%)
PC3D	88	80
SS	89	84
WR	93	80
RWR	100	92

TABLE II: Performance of robust grasping policies with known state (3D object shape and pose) across 75 physical trials per policy on the Adversarial object dataset. The policies differ by the metric used to rank grasps, and each metric is computed using the known 3D object geometry. The robust wrench resistance metric, which considers the ability of a suction cup to form a seal and resist gravity under perturbations, has very high precision. In comparison, the Planarity-Centroid heuristic achieves only 88% precision and 80% success.

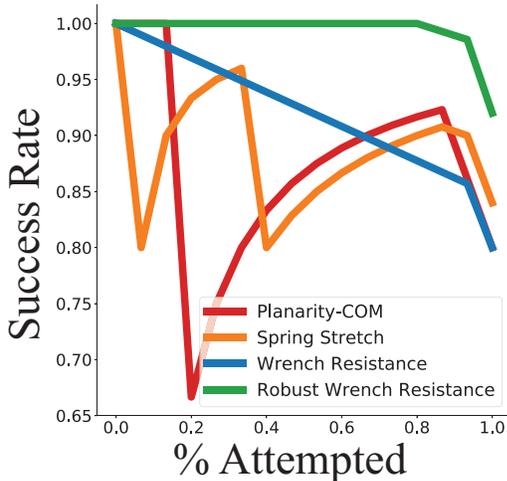


Fig. 7: Success rate vs Attempt Rate for grasp quality metrics on known 3D objects in known poses. The data was collected across 75 trials per policy on the Adversarial object dataset. The robust wrench resistance metric based on our compliant suction contact model had a 100% success rate for a large percentage of possible test cases, whereas a heuristic based on planarity and the distance to the object center of mass had success rates as low as 67%, indicating that the real-valued distance to the center of mass is not well correlated with grasp success.

and Typical classes. Also, the GQ-CNN trained on the Adversarial objects significantly outperforms all methods on the Adversarial dataset, suggesting that our model is able to exploit knowledge of complex 3D geometry to plan robust grasps. Furthermore, the Success-Attempt Rate curve suggests that the continuous probability of success output by the Dex-Net 3.0 policy is highly correlated with the true success label and can be used to detect failures before they occur on the Basic and Typical object classes.

D. Classification Performance on Known Objects

To assess performance of our robustness metric on classifying grasps as successful or unsuccessful, we evaluated whether or not the metric was able to classify a set of grasps sampled randomly from the 3D object surface using the known 3D object geometry and pose of the Adversarial objects. First, we sampled a set of 1000 grasps uniformly

Metric	AP (%)	Accuracy (%)	Rank Correlation
PC3D	71	68	0.36
SS	75	74	0.49
WR	78	77	0.52
RWR	80	75	0.62

TABLE III: Performance of classification and correlation with successful object lifts and transports for various metrics of grasp quality based on 3D object meshes. The metrics SS, WR, and RWR all use our compliant suction contact model, and RWR uses our entire proposed method: checking seal formation, analyzing wrench resistance using the suction ring model, and computing robustness with Monte-Carlo sampling.

at random from the surface of the 3D object meshes. Then robot was presented one of the five Adversarial objects in a known stable pose, selected from the top three most probable stable poses. We hand-aligned the object to a template image generated by rendering the object in a known pose on the table. Then, the robot executed a grasp uniformly at random from the set of reachable grasps for the given stable pose. In total, there were 600 trials, 125 per object.

We compared the predictions made for those grasps by the following metrics:

- 1) **Planarity-Centroid (PC3D)**. The inverse distance to the object centroid for sufficiently planar patches on the 3D object surface.
- 2) **Spring Stretch (SS)**. The maximum stretch among virtual springs in the suction contact model.
- 3) **Wrench Resistance (WR)**.
- 4) **Robust Wrench Resistance (RWR)**.

We measured the Average Precision (AP), classification accuracy, and Spearman’s rank correlation coefficient (which measures the correlation between the ranking of the metric value and successes on the physical robot). Table III details the performance of each metric and the Precision-Recall curve is plotted in Fig. 9. We see that the robust wrench resistance metric with our compliant spring contact model has the highest average precision and correlation with successes on the physical robot.

E. Failure Modes

Our system was not able to handle many objects due to material properties. We broke up the failure objects into two categories:

- 1) **Imperceptible Objects**: Objects with (a) surface variations less than the spatial resolution of our Prime-sense Carmine 1.09 depth camera or (b) specularities / transparencies that prevent the depth camera from sensing the object geometry. Thus the point-cloud-based grasping policies were not able to distinguish successes from failures.
- 2) **Impossible Objects**: Objects for which a seal cannot be formed either because objects are (a) non-porous or (b) lack a surface patch for which the suction cup can achieve a seal due to size or texture.

These objects are illustrated in Fig. 10.

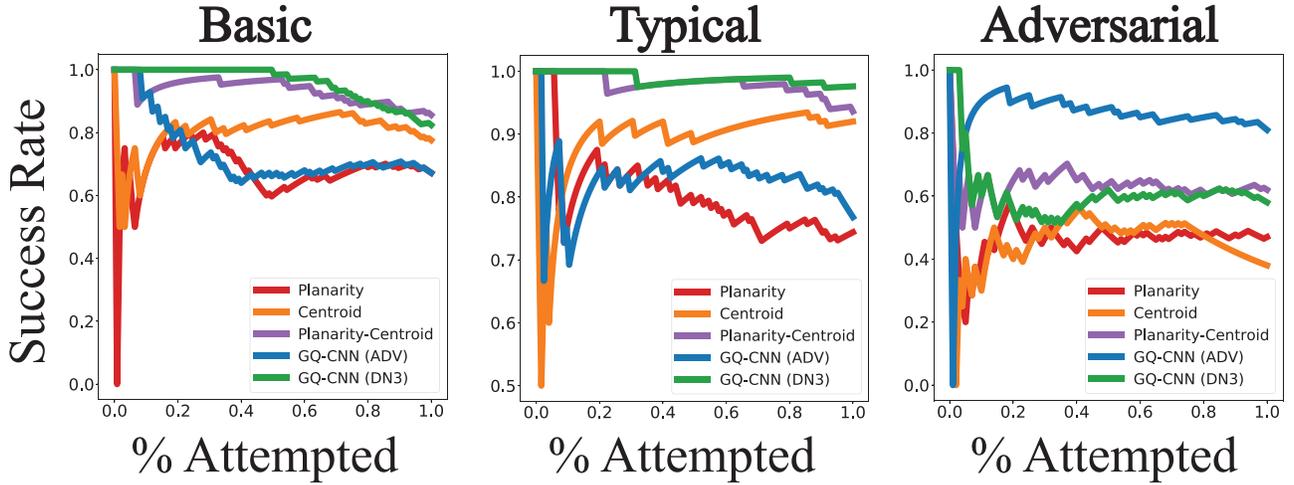


Fig. 8: Success vs Attempt Rate for 125 trials on each of the Basic and Typical object datasets and 100 trials each on the Adversarial object dataset. The GQ-CNN trained on Dex-Net 3.0 has near 100% precision on the Basic and Typical classes for a significant portion of attempts, suggesting that the GQ-CNN is able to predict when it is likely to fail on novel objects. The GQ-CNN trained on the Adversarial objects has a significantly higher precision on the Adversarial class but does not perform as well on the other objects.

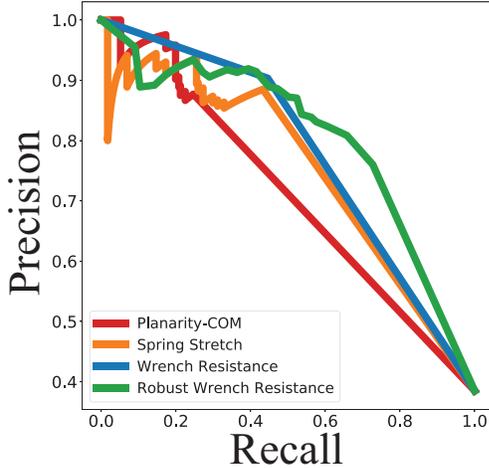


Fig. 9: Precision-Recall curve for classifying successful object lifts and transports using various metrics of grasp quality based on 3D object meshes.



Fig. 10: Two categories that cannot be handled by any of the point-cloud based suction grasping policies. **(Left)** Imperceptible objects, which cannot be handled by the system due to small surface variations that cannot be detected by the low-resolution depth sensor but do prevent seal formation. **(Right)** Impossible objects, which cannot be handled by the system due to non-porosity or lack of an available surface to form a seal.

APPENDIX II

DETAILS OF QUASI-STATIC SPRING SEAL FORMATION MODEL

In this section, we derive a detailed process for statically determining a final configuration of C that achieves a com-

plete seal against mesh M . We assume that we are given a line of approach ℓ parameterized by \mathbf{p} , a target point on the surface of M , and \mathbf{v} , a vector pointing towards P along the line of approach.

First, we choose an initial, undeformed configuration of C . In this undeformed configuration of C , all of the springs of C are in their resting positions, which means that the structural springs of C form a right pyramid with a regular n -gon as its base. This perfectly constrains the relative positions of the vertices of C , so all that remains is specifying the position and orientation of C relative to the world frame.

We further constrain the position and orientation of C such that ℓ passes through \mathbf{a} and is orthogonal to the plane containing the base of C . This leaves only the position of \mathbf{a} and a rotation about ℓ as degrees of freedom. For our purposes, the position of \mathbf{a} along ℓ does not matter so long as C is not in collision with M and the base of C is closer to M than the apex is. In general, we choose \mathbf{a} such that $\|\mathbf{p} - \mathbf{a}\| > x + h$, where x is the largest extent of the object's vertices. For the rotation about ℓ , we simply select a random initial angle. Theoretically, the rotation could affect the outcome of our metric, but as long as n is chosen to be sufficiently large, the result is not sensitive to the chosen rotation angle.

Next, given the initial configuration of C , we compute the final locations of the perimeter springs on the surface of M under two main constraints:

- The perimeter springs of C must not deviate from their initial locations when projected back onto the plane containing the base of C 's initial right pyramid.
- The perimeter springs of C must lie flush against the mesh M .

Essentially, this means that the perimeter springs will lie on the intersection of M with a right prism K whose base is the base of the initial configuration's right pyramid and whose

height is sufficient such that K passes all the way through M . The base vertices of C will lie at the intersection of M and K 's side edges, and the perimeter springs of C will lie along the intersection of M and K 's side faces.

Finally, given a complete configuration of the perimeter vertices of C as well as the paths of the perimeter springs along the surface of M , we compute the final location of the cup apex \mathbf{a} . We work with three main constraints:

- \mathbf{a} must lie on ℓ .
- \mathbf{a} must not be below the surface of M (i.e. $\mathbf{v}^T(\mathbf{a}-\mathbf{p}) \leq 0$).
- \mathbf{a} should be chosen such that the average displacement between \mathbf{a} and the perimeter vertices along \mathbf{v} remains equal to h .

Let $\mathbf{a}^* = \mathbf{p} - t^*\mathbf{v}$. Then, the solution distance t^* is given by

$$t^* = \min \left(\left[\frac{1}{n} \sum_{i=1}^n (\mathbf{v}_i - \mathbf{p})^T \mathbf{v} \right] - h, 0 \right).$$

When thresholding the energy in each spring, we use a per-spring threshold of a 10% change in length, which was used as the spring stretch limit in [30].

APPENDIX III SUCTION CONTACT MODEL

The basis of contact wrenches for the suction ring model is illustrated in Fig. 11. The contact wrenches are not independent due to the coupling of normal force and friction, and they may be bounded due to material properties. In this section we prove that wrench resistance can be computed with quadratic programming, we derive constraints between the contact wrenches in the suction ring model, and we explain the limits of the soft finger suction contact models for a single suction contact.

A. Computing Wrench Resistance with Quadratic Programming

The object wrench set for a grasp using a contact model with m basis wrenches is $\Lambda = \{\mathbf{w} \in \mathbb{R}^6 \mid \mathbf{w} = G\alpha \text{ for some } \alpha \in \mathcal{F}\}$, where $G \in \mathbb{R}^{6 \times m}$ is a set of m basis wrenches in the object coordinate frame, and $\mathcal{F} \subseteq \mathbb{R}^m$ is a set of constraints on contact wrench magnitudes [28]. The grasp map G can be decomposed as $G = AW$ where $A \in \mathbb{R}^{6 \times 6}$ is the *adjoint transformation* mapping wrenches from the contact to the object coordinate frame and $W \in 6 \times m$ is the *contact wrench basis*, a set of m orthonormal basis wrenches in the contact coordinate frame [28].

Definition 4: A grasp \mathbf{u} achieves *wrench resistance* with respect to \mathbf{w} if $-\mathbf{w} \in \Lambda$.

Proposition 1: Let G be the grasp map for a grasp \mathbf{u} . Furthermore, let $\epsilon^* = \operatorname{argmin}_{\alpha \in \mathcal{F}} \|G\alpha + \mathbf{w}\|_2^2$. Then \mathbf{u} can resist \mathbf{w} iff $\epsilon^* = 0$.

Proof: \Rightarrow . Assume \mathbf{u} can resist \mathbf{w} . Then $-\mathbf{w} \in \Lambda$ and therefore $\exists \alpha \in \mathcal{F}$ such that $G\alpha = -\mathbf{w} \Rightarrow G\alpha + \mathbf{w} = \mathbf{0}$. \Leftarrow . Assume $\epsilon^* = 0$. Then $\exists \alpha \in \mathcal{F}$ such that $G\alpha + \mathbf{w} = \mathbf{0} \Rightarrow G\alpha = -\mathbf{w} \Rightarrow -\mathbf{w} \in \Lambda$. ■

When the set of admissible contact wrench magnitudes \mathcal{F} is defined by linear equality and inequality constraints, the

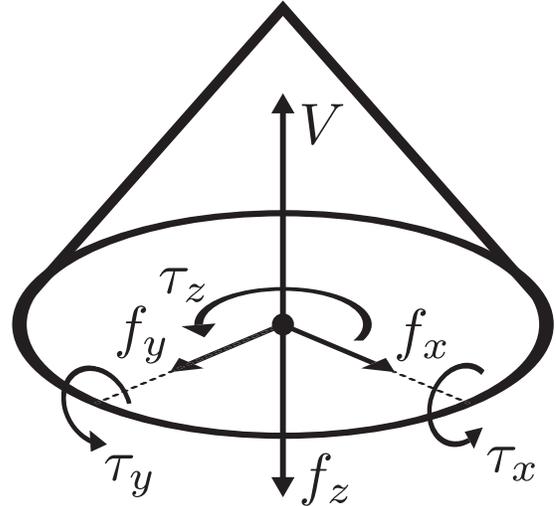


Fig. 11: Wrench basis for the compliant suction ring contact model. The contact exerts a constant pulling force on the object of magnitude V and additionally can push or pull the object along the contact z axis with force f_z . The suction cup material exerts a normal force $f_N = f_z + V$ on the object through a linear pressure distribution on the ring. This pressure distribution induces a friction limit surface bounding the set of possible frictional forces in the tangent plane $f_t = (f_x, f_y)$ and the torsional moment τ_z , and also induces torques τ_x and τ_y about the contact x and y axes due to elastic restoring forces in the suction cup material.

$\min_{\alpha \in \mathcal{F}} \|G\alpha + \mathbf{w}\|_2^2$, is a Quadratic Program which can be solved exactly by modern solvers.

B. Derivation of Suction Ring Contact Model Constraints

Our suction contact model assumes the following:

- 1) Quasi-static physics (e.g. inertial terms are negligible).
- 2) The suction cup contacts the object along a circle of radius r (or “ring”) in the xy -plane of the contact coordinate frame.
- 3) The suction cup material behaves as a ring of infinitesimal springs per unit length. Specifically, we assume that the pressure along the z -axis in contact coordinate frame satisfies $p(\theta) = k\delta_z(\theta)$ where δ_z is displacement along the z -axis and $k \in \mathbb{R}$ is a spring constant (per unit length). The cup does not permit deformations along the x or y axes.
- 4) The suction cup material is well approximated by a spring-mass system. Furthermore, points on the contact ring are in static equilibrium with a linear displacement along the z -axis from the equilibrium position: $\delta_z(\theta) = \delta_0 + ar \cos(\theta) + br \sin(\theta)$. Together with Assumption 3, this implies that:

$$p(\theta) = p_0 + p_x \cos(\theta) + p_y \sin(\theta)$$

for real numbers p_0, p_x , and p_y .

- 5) The force on the object due to the vacuum is a constant $-V$ along the z -axis of the contact coordinate frame.
- 6) The object exerts a normal force on the object $f_N = f_z + V$ where f_z is the force due to actuation. This assumption holds when analyzing the ability to resist disturbing wrenches because the material can apply

passive forces but may not hold when considering target wrenches that can be actuated.

The magnitudes of the contact wrenches are constrained due to (a) the friction limit surface [16], (b) limits on the elastic behavior of the suction cup material, and (c) limits on the vacuum force.

1) *Friction Limit Surface:* The values of the tangential and torsional friction are coupled through the planar external wrench and thus are jointly constrained. This constraint is known as the *friction limit surface* [16]. We can approximate the friction limit surface by computing the maximum friction force and torsional moment under a pure translation and a pure rotation about the contact origin.

The tangential forces have maximum magnitude under a purely translational disturbing wrench with unit vector $\hat{\mathbf{v}}$ in the direction of the velocity:

$$\begin{aligned} f_x &\leq \int_0^{2\pi} \mu \hat{\mathbf{v}}_x p(\theta) d\theta \\ &\leq \int_0^{2\pi} \mu \hat{\mathbf{v}}_x (p_0 + p_x \cos(\theta) + p_y \sin(\theta)) d\theta \\ &\leq 2\pi \mu \hat{\mathbf{v}}_x p_0 \\ f_y &\leq 2\pi \mu \hat{\mathbf{v}}_y p_0 \\ \|f_f\|_2^2 &\leq (2\pi \mu \hat{\mathbf{v}}_x p_0)^2 + (2\pi \mu \hat{\mathbf{v}}_y p_0)^2 \\ &= (2\pi \mu p_0)^2 (\hat{\mathbf{v}}_x^2 + \hat{\mathbf{v}}_y^2) \\ &= (2\pi \mu p_0)^2 \\ &= \mu f_N \end{aligned}$$

The torsional moment has a maximum moment under a purely rotational disturbing wrench about the contact z -axis. This disturbing wrench can be described with a unit vector $\hat{\mathbf{v}}(\theta) = (\sin(\theta), -\cos(\theta), 0)$. Thus the torsional moment is bounded by:

$$\begin{aligned} |\tau_z| &\leq \int_0^{2\pi} \mu r p(\theta) d\theta = \int_0^{2\pi} \mu r (p_0 + p_x \cos(\theta) + p_y \sin(\theta)) d\theta \\ &\leq 2\pi \mu r p_0 \\ &\leq r \mu f_N \end{aligned}$$

We can approximate the friction limit surface by the ellipsoid [15], [16]:

$$\frac{\|f_t\|_2^2}{(\mu f_N)^2} + \frac{|\tau_z|^2}{(r \mu f_N)^2} \leq 1$$

While this constraint is convex, in practice many solvers for Quadratically Constrained Quadratic Programs (QCQPs) assume nonconvexity. We can turn this into a linear constraint by bounding tangential forces and torsional moments in a

rectangular prism inscribed within the ellipsoid:

$$\begin{aligned} |f_x| &\leq \frac{\sqrt{3}}{3} \mu f_N \\ |f_y| &\leq \frac{\sqrt{3}}{3} \mu f_N \\ |\tau_z| &\leq \frac{\sqrt{3}}{3} r \mu f_N \end{aligned}$$

2) *Elastic Restoring Torques:* The torques about the x and y axes are also bounded. Let $\mathbf{w}(\theta) = (r \cos(\theta), r \sin(\theta), 0)$. Then:

$$\begin{aligned} \tau_t &= \int_0^{2\pi} (\mathbf{w}(\theta) \times \mathbf{e}_z) p(\theta) d\theta \\ \tau_x &= \int_0^{2\pi} r \sin(\theta) p(\theta) d\theta \\ &= \int_0^{2\pi} r \sin(\theta) (p_0 + p_x \cos(\theta) + p_y \sin(\theta)) d\theta \\ &= \int_0^{2\pi} r p_y \sin^2(\theta) d\theta \\ &= \pi r p_y \\ \tau_y &= \pi r p_x \\ \|\tau_e\|_2^2 &= \pi^2 r^2 (p_x^2 + p_y^2) \\ &\leq \pi^2 r^2 \kappa^2 \end{aligned}$$

where κ is a material constant depending on the linear elasticity of the material.

3) *Vacuum Limits:* The ring contact can exert forces f_z on the object along the z axis through motor torques that transmit forces to the object through the ring of the suction cup. Under these assumptions, the normal force exerted on the object by the suction cup material is:

$$\begin{aligned} f_N &= \int_0^{2\pi} p(\theta) d\theta = \int_0^{2\pi} (p_0 + p_x \cos(\theta) + p_y \sin(\theta)) d\theta \\ &= 2\pi p_0 \end{aligned}$$

Note also that $f_N = f_z + V$, where f_z is the z component of force on the object, since the normal force must offset the force due to vacuum V even when no force is being applied on the object.

4) *Constraint Set:* Taking all constraints into account, we can describe \mathcal{F} with a set of linear constraints:

$$\begin{aligned} \mathbf{Friction:} \quad &\sqrt{3}|f_x| \leq \mu f_N \quad \sqrt{3}|f_y| \leq \mu f_N \quad \sqrt{3}|\tau_z| \leq r \mu f_N \\ \mathbf{Material:} \quad &\sqrt{2}|\tau_x| \leq \pi r \kappa \quad \sqrt{2}|\tau_y| \leq \pi r \kappa \\ \mathbf{Suction:} \quad &f_z \geq -V \end{aligned}$$

Since these constraints are linear, we can solve for wrench resistance in the our contact model using Quadratic Programming. In this paper we set $V = 250N$ and $\kappa = 0.005$.

C. Limits of the Soft Finger Suction Contact Model

The most common suction contact model in the literature [19], [26], [34], [36], [40] considers normal forces from motor torques, suction forces from the pressure differential between inside the cup and the air outside the object, and both tangential and torsional friction resulting from the contact area between the cup and the object. Let \mathbf{e}_x , \mathbf{e}_y , and \mathbf{e}_z be unit basis vectors along the x , y , and z axes. The contact model is specified by:

$$W = \begin{bmatrix} \mathbf{e}_x & \mathbf{e}_y & \mathbf{e}_z & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{e}_z \end{bmatrix}$$

$$\alpha = (f_x, f_y, f_z, \tau_z) \in \mathcal{F} \text{ if and only if:}$$

$$\sqrt{f_x^2 + f_y^2} \leq \mu f_z$$

$$|\tau_z| \leq \gamma |f_z|$$

The first constraint enforces Coulomb friction with coefficient μ . The second constraint ensures that the net torsion is bounded by the normal force, since torsion results from the net frictional moment from a contact area. Unlike contact models for rigid multifinger grasping, f_z can be positive or negative due to the pulling force of suction.

Proposition 2: Under the soft suction contact model, a grasp with a single contact point cannot resist torques about axes in the contact tangent plane.

Proof: The wrench $\mathbf{w} = (\mathbf{0}, \tau_e)$ is not in the range of W because it is orthogonal every basis wrench (column of W). ■

The null space of W is spanned by the wrenches $\mathbf{w}_1 = (\mathbf{0}, \mathbf{e}_x)$ and $\mathbf{w}_2 = (\mathbf{0}, \mathbf{e}_y)$, suggesting that a single suction contact cannot resist torques in the tangent plane at the contact. This defies our intuition since empirical evidence suggests that a single point of suction can reliably hold and transporting objects to a receptacle in applications such as the Amazon Picking Challenge [7], [12].

APPENDIX IV GQ-CNN TRAINING

The GQ-CNN trained on Dex-Net 3.0 had an accuracy of 93.5% on a held out validation set of approximately 552,000 datapoints. Fig. 12 shows the precision-recall curve for the GQ-CNN validation set and the optimized 64 Conv1.1 filters, each of which is 7×7 .

APPENDIX V ENVIRONMENT MODEL

To learn to predict grasp robustness based on noisy point clouds, we generate the Dex-Net 3.0 training dataset of point clouds, grasps, and grasp success labels by sampling tuples $(S_i, \mathbf{u}_i, \mathbf{y}_i)$ from a joint distribution $p(S, \mathbf{u}, \mathbf{x}, \mathbf{y})$ that is composed of distributions on:

- **States:** $p(\mathbf{x})$: A prior on possible objects, object poses, and camera poses that the robot will encounter.
- **Grasp Candidates:** $p(\mathbf{u}|\mathbf{x})$: A prior constraining grasp candidates to target points on the object surface.
- **Grasp Successes** $p(S | \mathbf{u}, \mathbf{x})$: A stochastic model of wrench resistance for the gravity wrench.

Distribution	Description
$p(\gamma)$	truncated Gaussian distribution over friction coefficients
$p(\mathcal{O})$	discrete uniform distribution over 3D object models
$p(T_o \mathcal{O})$	continuous uniform distribution over the discrete set of object stable poses and planar poses on the table surface
$p(T_c)$	continuous uniform distribution over spherical coordinates for radial bounds $[r_\ell, r_u]$ and polar angle in $[0, \delta]$

TABLE IV: Details of the distributions used in the Dex-Net 3.0 environment model for generating the Dex-Net 3.0 dataset. Source of table: [24]

- **Observations** $p(\mathbf{y} | \mathbf{x})$: A sensor noise model.

Our graphical model is illustrated in Fig. 13

A. Details of Distributions

We follow the state model of [24], which we repeat here for convenience. We model the state distribution as $p(\mathbf{x}) = p(\gamma)p(\mathcal{O})p(T_o|\mathcal{O})p(T_c)$. We model $p(\gamma)$ as a Gaussian distribution $\mathcal{N}(0.5, 0.1)$ truncated to $[0, 1]$. We model $p(\mathcal{O})$ as a discrete uniform distribution over 3D objects in a given dataset. We model $p(T_o|\mathcal{O}) = p(T_o|T_s)p(T_s|\mathcal{O})$, where $p(T_s|\mathcal{O})$ is a discrete uniform distribution over object stable poses and $p(T_o|T_s)$ is uniform distribution over 2D poses: $\mathcal{U}([-0.1, 0.1] \times [-0.1, 0.1] \times [0, 2\pi))$. We compute stable poses using the quasi-static algorithm given by Goldberg et al. [9]. We model $p(T_c)$ as a uniform distribution on spherical coordinates $r, \theta, \varphi \sim \mathcal{U}([0.5, 0.7] \times [0, 2\pi] \times [0.01\pi, 0.1\pi])$, where the camera optical axis always intersects the center of the table.

Our grasp candidate model $p(\mathbf{u} | \mathbf{x})$ is a uniform distribution over points samples on the object surface, with the approach direction defined by the inward-facing surface normal at each point.

We follow the observation model of [24], which we repeat here for convenience. Our observation model $p(\mathbf{y} | \mathbf{x})$ model images as $\mathbf{y} = \alpha * \hat{\mathbf{y}} + \epsilon$ where $\hat{\mathbf{y}}$ is a rendered depth image created using OSMesa offscreen rendering. We model α as a Gamma random variable with shape=1000.0 and scale=0.001. We model ϵ as Gaussian Process noise drawn with measurement noise $\sigma = 0.005$ and kernel bandwidth $\ell = \sqrt{2}px$.

Our grasp success model $p(S | \mathbf{u}, \mathbf{x})$ specifies a distribution over wrench resistance due to perturbations in object pose, gripper pose, friction coefficient, and the disturbing wrench to resist. Specifically, we model $p(S | \mathbf{u}, \mathbf{x}) = p(S | \hat{\mathbf{u}}, \hat{\mathbf{x}}, \mathbf{w})p(\hat{\mathbf{x}} | \mathbf{x})p(\hat{\mathbf{u}} | \mathbf{u})p(\mathbf{w})$. We model $p(\mathbf{w})$ as the wrench exerted by gravity on the object center-of-mass with zero-mean Gaussian noise $\mathcal{N}(\mathbf{0}_3, 0.01\mathbf{I}_3)$ assuming as mass of 1.0kg. We model $p(\hat{\mathbf{u}} | \mathbf{u})$ as a grasp perturbation distribution where the suction target point is perturbed by zero-mean Gaussian noise $\mathcal{N}(\mathbf{0}_3, 0.001\mathbf{I}_3)$ and the approach direction is perturbed by zero-mean Gaussian noise in the rotational component of Lie algebra coordinates $\mathcal{N}(\mathbf{0}_3, 0.1\mathbf{I}_3)$. We model $p(\hat{\mathbf{x}} | \mathbf{x})$ as a state perturbation distribution where the pose T_o is perturbed by zero-mean Gaussian noise in Lie algebra coordinates with translational component $\mathcal{N}(\mathbf{0}_3, 0.001\mathbf{I}_3)$ and rotational component $\mathcal{N}(\mathbf{0}_3, 0.1\mathbf{I}_3)$ and the object center of mass is perturbed by zero-mean Gaussian

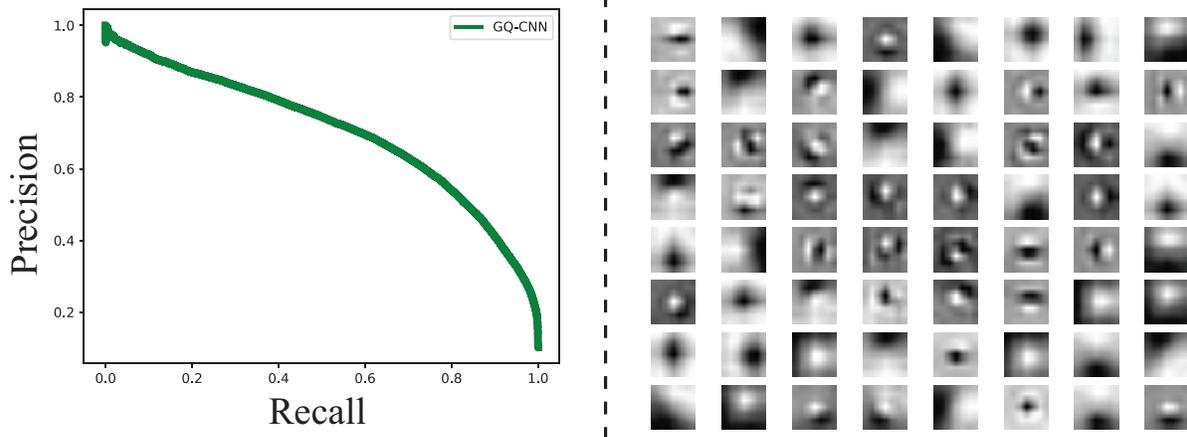


Fig. 12: **(Left)** Precision-recall curve for the GQ-CNN trained on Dex-Net 3.0 on the validation set of 552,000 pairs of grasps and images. **(Right)** The 64 Conv1.1 filters of the GQ-CNN. Each is 7×7 . We see that the network learns circular filters which may be used to assess the surface curvature about the ring of contact between the suction cup and object.

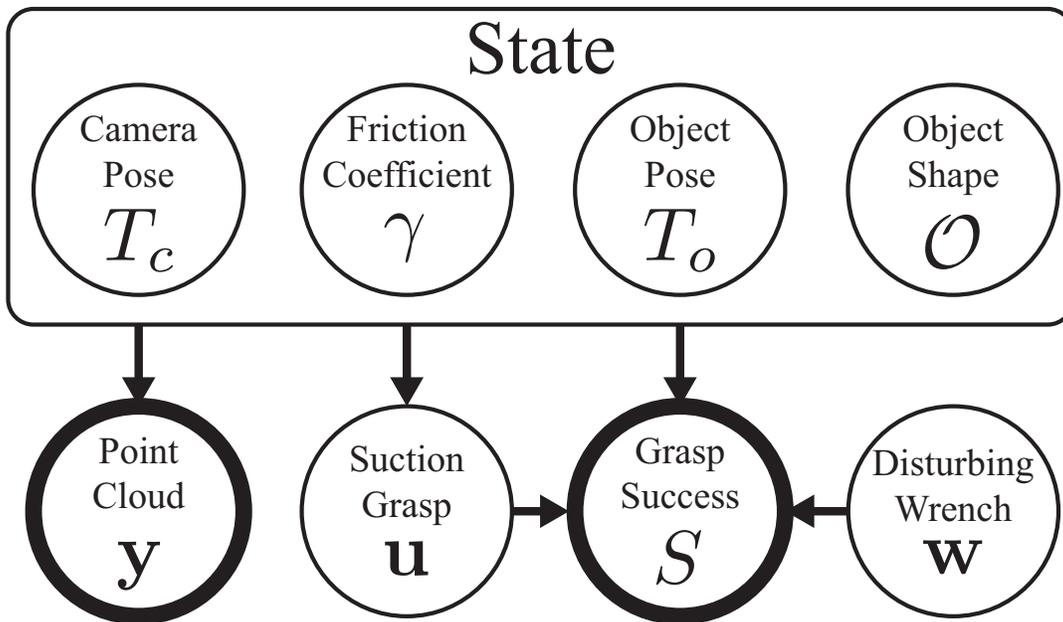


Fig. 13: A probabilistic graphical model of the relationship between the ability to resist external wrenches e.g. due to gravity under perturbations in object pose, gripper pose, camera pose, and friction.

noise $\mathcal{N}(\mathbf{0}_3, 0.0025\mathbf{I}_3)$. We model $p(S | \hat{\mathbf{u}}, \hat{\mathbf{x}}, \mathbf{w})$ a Bernoulli with parameter 1 if $\hat{\mathbf{u}}$ resists \mathbf{w} given the state $\hat{\mathbf{x}}$ and parameter 0 if not.

B. Implementation Details

To efficiently implement sampling, we make several optimizations. First, we precompute the set of grasps for every 3D object model in the database and take a fixed number of samples of grasp success from $p(S | \mathbf{u}, \mathbf{x})$ using quadratic programming for wrench resistance evaluation. We convert the samples to binary success labels by thresholding the sample mean by $\tau = 0.5$. We also render a fixed number of depth images for each stable pose independently of grasp success evaluation. Finally, we sample a set of candidate grasps from the object in each depth image and transform

the image to generate a suction grasp thumbnail centered on the target point and oriented to align the approach axis with the middle column of pixels for GQ-CNN training.

REFERENCES

- [1] A. Ali, M. Hosseini, and B. Sahari, "A review of constitutive models for rubber-like materials," *American Journal of Engineering and Applied Sciences*, vol. 3, no. 1, pp. 232–239, 2010.
- [2] B. Bahr, Y. Li, and M. Najafi, "Design and suction cup analysis of a wall climbing robot," *Computers & electrical engineering*, vol. 22, no. 3, pp. 193–209, 1996.
- [3] J. Bohg, A. Morales, T. Asfour, and D. Kragic, "Data-driven grasp synthesis: a survey," *IEEE Trans. Robotics*, vol. 30, no. 2, pp. 289–309, 2014.
- [4] N. Correll, K. E. Bekris, D. Berenson, O. Brock, A. Causo, K. Hauser, K. Okada, A. Rodriguez, J. M. Romano, and P. R. Wurman, "Analysis and observations from the first amazon picking challenge," *IEEE Transactions on Automation Science and Engineering*, 2016.

- [5] R. Detry, C. H. Ek, M. Madry, and D. Kragic, "Learning a dictionary of prototypical grasp-predicting parts from grasping experience," in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA)*. IEEE, 2013, pp. 601–608.
- [6] Y. Domae, H. Okuda, Y. Taguchi, K. Sumi, and T. Hirai, "Fast graspability evaluation on single depth maps for bin picking with general grippers," in *Robotics and Automation (ICRA), 2014 IEEE International Conference on*. IEEE, 2014, pp. 1997–2004.
- [7] C. Eppner, S. Höfer, R. Jonschkowski, R. M. Martin, A. Sieverling, V. Wall, and O. Brock, "Lessons from the amazon picking challenge: Four aspects of building robotic systems," in *Robotics: Science and Systems*, 2016.
- [8] C. Ferrari and J. Canny, "Planning optimal grasps," in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA)*, 1992, pp. 2290–2295.
- [9] K. Goldberg, B. V. Mirtich, Y. Zhuang, J. Craig, B. R. Carlisle, and J. Canny, "Part pose statistics: Estimators and experiments," *IEEE Trans. Robotics and Automation*, vol. 15, no. 5, pp. 849–857, 1999.
- [10] M. Haas-Heger, G. Iyengar, and M. Ciocarlie, "On the distinction between active and passive reaction in grasp stability analysis," in *Workshop on the Algorithmic Foundation of Robotics (WAFR)*, 2016.
- [11] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003.
- [12] C. Hernandez, M. Bharatheesha, W. Ko, H. Gaiser, J. Tan, K. van Deurzen, M. de Vries, B. Van Mil, J. van Egmond, R. Burger *et al.*, "Team delft's robot winner of the amazon picking challenge 2016," *arXiv preprint arXiv:1610.05514*, 2016.
- [13] M. Jaderberg, K. Simonyan, A. Zisserman *et al.*, "Spatial transformer networks," in *Advances in Neural Information Processing Systems*, 2015, pp. 2017–2025.
- [14] E. Johns, S. Leutenegger, and A. J. Davison, "Deep learning a grasp function for grasping under gripper pose uncertainty," in *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*. IEEE, 2016, pp. 4461–4468.
- [15] I. Kao and M. R. Cutkosky, "Quasistatic manipulation with compliance and sliding," *Int. Journal of Robotics Research (IJRR)*, vol. 11, no. 1, pp. 20–40, 1992.
- [16] I. Kao, K. Lynch, and J. W. Burdick, "Contact modeling and manipulation," in *Springer Handbook of Robotics*. Springer, 2008, pp. 647–669.
- [17] D. Kappler, J. Bohg, and S. Schaal, "Leveraging big data for grasp planning," in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA)*, 2015.
- [18] A. Kasper, Z. Xue, and R. Dillmann, "The kit object models database: An object model database for object recognition, localization and manipulation in service robotics," *Int. Journal of Robotics Research (IJRR)*, vol. 31, no. 8, pp. 927–934, 2012.
- [19] R. Kolluru, K. P. Valavanis, and T. M. Hebert, "Modeling, analysis, and performance evaluation of a robotic gripper system for limp material handling," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 28, no. 3, pp. 480–486, 1998.
- [20] R. Krug, Y. Bekiroglu, and M. A. Roa, "Grasp quality evaluation done right: How assumed contact force bounds affect wrench-based quality metrics," in *Robotics and Automation (ICRA), 2017 IEEE International Conference on*. IEEE, 2017, pp. 1595–1600.
- [21] I. Lenz, H. Lee, and A. Saxena, "Deep learning for detecting robotic grasps," *Int. Journal of Robotics Research (IJRR)*, vol. 34, no. 4-5, pp. 705–724, 2015.
- [22] S. Levine, P. Pastor, A. Krizhevsky, and D. Quillen, "Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection," *arXiv preprint arXiv:1603.02199*, 2016.
- [23] Z. Li and S. S. Sastry, "Task-oriented optimal grasping by multifingered robot hands," *IEEE Journal on Robotics and Automation*, vol. 4, no. 1, pp. 32–44, 1988.
- [24] J. Mahler, J. Liang, S. Niyaz, M. Laskey, R. Doan, X. Liu, J. A. Ojeda, and K. Goldberg, "Dex-net 2.0: Deep learning to plan robust grasps with synthetic point clouds and analytic grasp metrics," in *Proc. Robotics: Science and Systems (RSS)*, 2017.
- [25] J. Mahler, F. T. Pokorny, B. Hou, M. Roderick, M. Laskey, M. Aubry, K. Kohlhoff, T. Kröger, J. Kuffner, and K. Goldberg, "Dex-net 1.0: A cloud-based network of 3d objects for robust grasp planning using a multi-armed bandit model with correlated rewards," in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA)*. IEEE, 2016.
- [26] G. Mantriota, "Theoretical model of the grasp with vacuum gripper," *Mechanism and machine theory*, vol. 42, no. 1, pp. 2–17, 2007.
- [27] M. Monta, N. Kondo, and K. Ting, "End-effectors for tomato harvesting robot," *Artificial Intelligence Review*, vol. 12, no. 1, pp. 11–25, 1998.
- [28] R. M. Murray, Z. Li, and S. S. Sastry, *A mathematical introduction to robotic manipulation*. CRC press, 1994.
- [29] L. Pinto and A. Gupta, "Supersizing self-supervision: Learning to grasp from 50k tries and 700 robot hours," in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA)*, 2016.
- [30] X. Provat *et al.*, "Deformation constraints in a mass-spring model to describe rigid cloth behaviour," in *Graphics interface*. Canadian Information Processing Society, 1995, pp. 147–147.
- [31] R. Y. Rubinstein, A. Ridder, and R. Vaisman, *Fast sequential Monte Carlo methods for counting and optimization*. John Wiley & Sons, 2013.
- [32] A. Saxena, J. Driemeyer, and A. Y. Ng, "Robotic grasping of novel objects using vision," *The International Journal of Robotics Research*, vol. 27, no. 2, pp. 157–173, 2008.
- [33] K. B. Shimoga, "Robot grasp synthesis algorithms: A survey," *The International Journal of Robotics Research*, vol. 15, no. 3, pp. 230–266, 1996.
- [34] H. S. Stuart, M. Bagheri, S. Wang, H. Barnard, A. L. Sheng, M. Jenkins, and M. R. Cutkosky, "Suction helps in a pinch: Improving underwater manipulation with gentle suction flow," in *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*. IEEE, 2015, pp. 2279–2284.
- [35] N. C. Tsourveloudis, R. Kolluru, K. P. Valavanis, and D. Gracanin, "Suction control of a robotic gripper: A neuro-fuzzy approach," *Journal of Intelligent & Robotic Systems*, vol. 27, no. 3, pp. 215–235, 2000.
- [36] A. J. Valencia, R. M. Idrovo, A. D. Sappa, D. P. Guingla, and D. Ochoa, "A 3d vision based approach for optimal grasp of vacuum grippers," in *Electronics, Control, Measurement, Signals and their Application to Mechatronics (ECMSM), 2017 IEEE International Workshop of*. IEEE, 2017, pp. 1–6.
- [37] J. Weisz and P. K. Allen, "Pose error robust grasping from contact wrench space metrics," in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA)*. IEEE, 2012, pp. 557–562.
- [38] W. Wohlkinger, A. Aldoma, R. B. Rusu, and M. Vincze, "3dnet: Large-scale object class recognition from cad models," in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA)*. IEEE, 2012, pp. 5384–5391.
- [39] L. Xin, W. Zhong, T. Kagawa, H. Liu, and G. Tao, "Development of a pneumatic sucker for gripping workpieces with rough surface," *IEEE Transactions on Automation Science and Engineering*, vol. 13, no. 2, pp. 639–646, 2016.
- [40] Y. Yoshida and S. Ma, "Design of a wall-climbing robot with passive suction cups," in *Robotics and Biomimetics (ROBIO), 2010 IEEE International Conference on*. IEEE, 2010, pp. 1513–1518.
- [41] K.-T. Yu, N. Fazeli, N. Chavan-Dafle, O. Taylor, E. Donlon, G. D. Lankenau, and A. Rodriguez, "A summary of team mit's approach to the amazon picking challenge 2015," *arXiv preprint arXiv:1604.03639*, 2016.