2018 IEEE International Conference on Robotics and Automation (ICRA)
May 21-25, 2018, Brisbane, Australia

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

Jeffrey Mahler[1], Matthew Matl[1], Xinyu Liu[1], Albert Li[1], David Gealy[1], Ken Goldberg[1,2]

***Abstract*— Vacuum-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. Suction grasp planners often target planar surfaces on point clouds near the estimated centroid of an object. In this paper, we propose a compliant suction contact model that computes the quality of the seal between the suction cup and local target surface and a measure of the ability of the suction grasp to resist an external gravity wrench. To characterize grasps, we estimate robustness to perturbations in end-effector and object pose, material properties, and external wrenches. We analyze grasps across 1,500 3D object models to generate Dex-Net 3.0, a dataset of 2.8 million point clouds, suction grasps, and grasp robustness labels. We use Dex-Net 3.0 to train a Grasp Quality Convolutional Neural Network (GQ-CNN) to classify robust suction targets in point clouds containing a single object. We evaluate the resulting system in 350 physical trials on an ABB YuMi fitted with a pneumatic suction gripper. When evaluated on novel objects that we categorize as Basic (prismatic or cylindrical), Typical (more complex geometry), and Adversarial (with few available suction-grasp points) Dex-Net 3.0 achieves success rates of 98%, 82%, and 58% respectively, improving to 81% in the latter case when the training set includes only adversarial objects. Code, datasets, and supplemental material can be found at http://berkeleyautomation.github.io/dex-net.**

I. Introduction

Suction grasping is widely-used 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 has been published on planning suction grasps.

While grasp planning searches for gripper configurations that maximize a quality metric derived from mechanical wrench space analysis [24], human labels [28], or self-supervised labels [19], suction grasps are often planned directly on point clouds using heuristics such as grasping near the object centroid [10] or at the center of planar surfaces [4], [5]. These heuristics work well for prismatic objects such as boxes and cylinders but may fail on objects with non-planar surfaces near the object centroid, which is common for industrial parts and household objects such as staplers or children's toys. Analytic models of suction cups for grasp planning 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 is perfectly known [2], [16], [22]. 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.

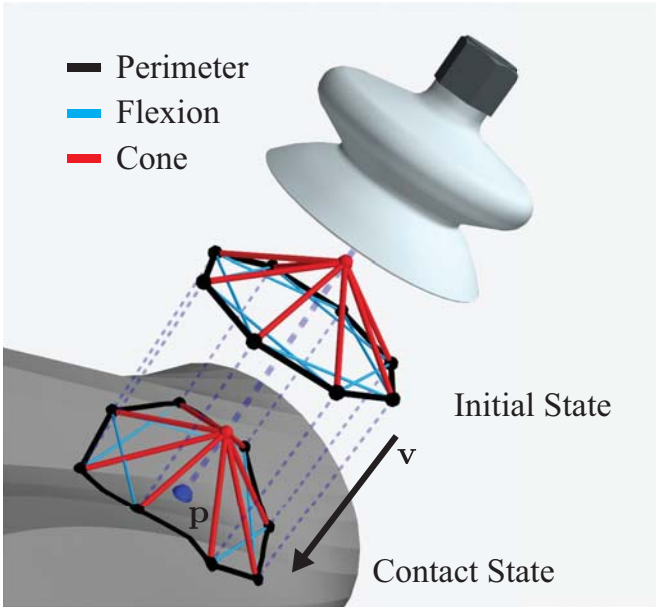

Fig. 1: The quasi-static spring model, C, used for determining when seal formation is feasible. The model contains three types of springs – perimeter, flexion, and cone springs. An initial state for C is chosen given a target point $\mathbf{p}$ and an approach direction $\mathbf{v}$. Then, a contact 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.

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 surface and (2) an analysis of the ability of the suction contact to resist external wrenches. We use the model to evaluate grasp robustness by analyzing 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 [21].

This paper makes four contributions:

1) A compliant suction contact model that quantifies seal formation using a quasi-static spring system and 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.

[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, alberthli, dgealy, goldberg}@berkeley.edu

2) Robust wrench resistance: a robust version of the above model under random disturbing wrenches and 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) Physical robot experiments measuring the precision of robust wrench resistance both with and without knowledge of the target object's shape and pose.

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 with several heuristics such as targeting planar surfaces near object centroids. We find that the method achieves success rates of 98%, 82%, and 58% on datasets of Basic (prismatic or cylindrical), Typical (more complex geometry), and Adversarial (with few available suction-grasp points), 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 [16], and robotics applications such as the Amazon Picking Challenge [4], underwater manipulation [29], 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. [26] 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 [26] 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 [16], [29], [31]. Bahr et al. [2] augmented this model with the ability 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 [22] modeled torsional friction due to a contact area between the cup and object similar to the soft finger contact model used in grasping [13]. Our model extends these methods by combining models of torsional friction [22] and contact moments [2] in a compliant model of the ring of contact between the cup and object.

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], which can be arbitrary [7] or task-specific [17]. A common approach is to select a configuration that maximizes a quality metric (or reward) based on wrench space analysis [24], robustness to perturbations [32], or a model learned from human labels [14] or self-supervision [25].

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 [16], [29], [30], [31]. Domae et al. [5] developed a geometric model to evaluate suction success by convolving target locations in images with a desired suction contact template to assess planarity. Heuristics for planning suction grasps from point clouds have also been used extensively in the Amazon Robotics Challenge. In 2015, Team RBO [6] won by pushing objects from the top or side until suction was achieved, and Team MIT [34] came in second place by suctioning on the centroid of objects with flat surfaces. In 2016, Team Delft [10] won the challenge by approaching the estimated object centroid along the inward surface normal. In 2017, Cartman [23] won the challenge and ranked suction grasps according to heuristics such as maximizing distance to the segmented object boundary and MIT [35] performed well using a neural network trained to predict grasp affordance maps from human labeled RGB-D point clouds. In this work, we present a novel metric that evaluates whether a single suction cup can resist external wrenches under perturbations in object / gripper poses, friction coefficient, 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 [18] or point clouds [14] 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 [19], [25]. To reduce the time-cost of data collection, recent research has proposed to generate labels in simulation using physical models of contact [12], [14]. Mahler et al. [20] 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 parallel-jaw 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 an object in isolation on a planar

worksurface and transport it to a receptacle. We compute the suction grasp that maximizes the probability that the robot can hold the object under gravity and perturbations sampled from a distribution over sensor noise, control imprecision, and random disturbing wrenches.

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 [8].
4) A single overhead depth sensor 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 reward function** R, where $R = 1$ if the grasp $\mathbf{u}$ successfully transports the object, and $R = 0$ 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 **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 $\mathbb{P}(R \mid \mathbf{x}, \mathbf{u})$, the ability of a grasp to resist random disturbing wrenches, and an observation model $p(\mathbf{y} \mid \mathbf{x})$. 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 $\mathbf{y}$ is the probability of grasp success under uncertainty in sensing, control, and disturbing wrenches: $Q(\mathbf{y}, \mathbf{u}) = \mathbb{P}(R \mid \mathbf{y}, \mathbf{u})$.

Our environment model is described in Section V and further details are given in the supplemental file.

C. Objective

Our ultimate goal is to find a grasp that maximizes robustness given a point cloud, $\pi^*(\mathbf{y}) = \text{argmax}_{\mathbf{u} \in \mathcal{C}} Q(\mathbf{y}, \mathbf{u})$, where $\mathcal{C}$ specifies constraints on the set of available grasps, such as collisions or kinematic feasibility. We approximate π^* by optimizing the weights θ of a deep Grasp Quality Convolutional Neural Network (GQ-CNN) $Q_\theta(\mathbf{y}, \mathbf{u})$ on a training dataset $\mathcal{D} = \{(\mathbf{y}_i, \mathbf{u}_i, R_i)\}_{i=1}^N$ consisting of reward values, point clouds, and suction grasps sampled from our stochastic model of grasp success. Our optimization objective is to find weights θ that minimize the cross-entropy loss $\mathcal{L}$ over $\mathcal{D}$:

$$\theta^* = \underset{\theta \in \Theta}{\text{argmin}} \sum_{i=1}^{N} \mathcal{L}(R_i, Q_\theta(\mathbf{y}_i, \mathbf{u}_i)). \tag{III.1}$$

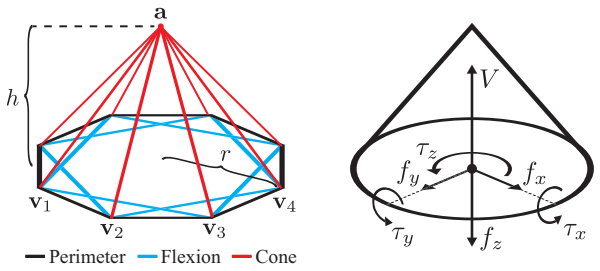

Fig. 2: Our compliant suction contact model. **(Left)** The quasi-static spring model used in seal formation computations. This suction cup is approximated by $n = 8$ components. Here, r is equal to the radius of the cup and h is equal to the height of the cup. $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ 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.

IV. Compliant Suction Contact Model

To quantify grasp robustness, we develop a quasi-static spring model of the suction cup material and a model of contact wrenches that the suction cup can apply to the object through a ring of contact on the suction cup perimeter. Under our model, the reward $R = 1$ if:

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 can lift objects due to an air pressure differential induced across the membrane of the cup by a vacuum generator that forces the object into the cup. 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, a tight seal between the cup and the target object is important for achieving a successful grasp.

To determine when seal formation is possible, we model circular suction cups as a conical spring system $\mathcal{C}$ parameterized by real numbers (n, r, h), where n is the numer of vertices along the contact ring, r is the radius of the cup, and h is the height of the cup. See see Fig. 2 for an illustration.

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 $\mathcal{C}$ that achieves a seal by projecting $\mathcal{C}$ onto the surface of the target object's triangular mesh $\mathcal{M}$ and evaluate the feasibility of that configuration under quasi-static conditions as a proxy for the dynamic feasibility of seal formation.

In our model, $\mathcal{C}$ 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 structures have been used in prior work to model stiff sheets of rubber [26]. The undeformed structural springs of $\mathcal{C}$ form a right pyramid with height h and with a base that is a regular n-gon with circumradius r. Let $\mathcal{V} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n, \mathbf{a}\}$ 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 $\mathbf{v}_{i+1}$, $\forall i \in \{1, \ldots, n\}$.
- **Cone (Structural) Springs** – Springs linking vertex $\mathbf{v}_i$ to vertex $\mathbf{a}$, $\forall i \in \{1, \ldots, n\}$.
- **Flexion Springs** – Springs linking vertex $\mathbf{v}_i$ to vertex $\mathbf{v}_{i+2}$, $\forall i \in \{1, \ldots, n\}$.

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

- **The perimeter springs** of $\mathcal{C}$ 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 $\mathcal{M}$.
- **The apex**, $\mathbf{a}$, of $\mathcal{C}$ must lie on the approach line $(\mathbf{p}, \mathbf{v})$ and, given the locations of $\mathcal{C}$'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 the supplemental file for additional details.

Given this configuration, a seal is feasible if:

- The cone faces of $\mathcal{C}$ do not collide with $\mathcal{M}$ during approach or in the contact configuration.
- The surface of $\mathcal{M}$ has no holes within the contact ring traced out by $\mathcal{C}$'s perimeter springs.
- The energy required in each spring to maintain the contact configuration of $\mathcal{C}$ is below a real-valued threshold E modeling the maximum deformation of the suction cup material against the object surface.

We threshold the energy in individual springs rather than the total energy for $\mathcal{C}$ because air gaps are usually caused by local geometry.

B. Wrench Space Analysis

To determine the degree to which the suction cup can resist external wrenches such as gravity, 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 basis wrenches in the object coordinate frame, and $\mathcal{F} \subseteq \mathbb{R}^m$ is a set of constraints on contact wrench magnitudes [24].

Definition 2: A grasp $\mathbf{u}$ achieves *wrench resistance* with respect to $\mathbf{w}$ if $-\mathbf{w} \in \Lambda$ [17], [24].

We encode wrench resistance as a binary variable W such that $W = 0$ 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], [16], [22], [29] similar to a point contact with friction or soft finger model [24]. 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 the supplementary material for a detailed proof). This defies our intuition since empirical evidence suggests that a single point of suction can robustly transport objects [6], [10].

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. 2 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 by pressing into the object along the contact 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 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 [13], (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 approximated by a set of linear constraints for efficient computation of wrench resistance:

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

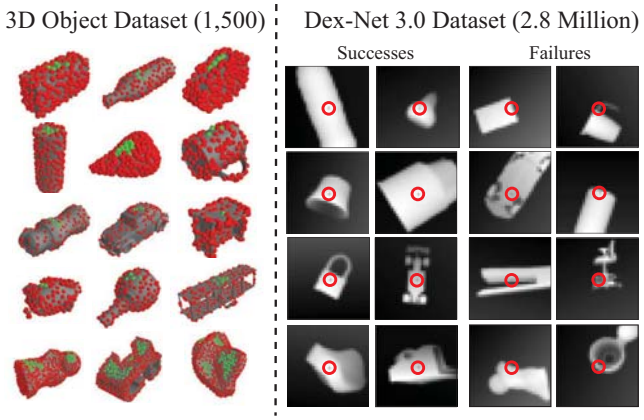

Fig. 3: The Dex-Net 3.0 dataset. (Left) The Dex-Net 3.0 object dataset contains approximately 350k unique suction target points across the surfaces of 1,500 3D models from the KIT object database [15] and 3DNet [33]. Each suction grasp is classified as robust (green) or non-robust (red). Robust grasps are often above the object center-of-mass on flat surfaces of the object. (Right) The Dex-Net 3.0 point cloud dataset contains 2.8 million tuples of point clouds and suction grasps with robustness labels, with approximately 11.8% positive examples.

Here μ is the friction coefficient, r is the radius of the contact ring, and κ is a material-dependent constant modeling the maximum stress for which the suction cup has linear-elastic behavior. 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 the supplemental file 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 $\lambda(\mathbf{u}, \mathbf{x}) = \mathbb{P}(W \mid \mathbf{u}, \mathbf{x})$, the probability of success under perturbations in object pose, gripper pose, friction, and disturbing wrenches.

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

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 reward labels by sampling tuples $(R_i, \mathbf{u}_i, \mathbf{y}_i)$ from a joint distribution $p(R, \mathbf{x}, \mathbf{y}, \mathbf{u})$ modeled as the product of distributions on:

- **States:** $p(\mathbf{x})$**:** A uniform distribution over a discrete dataset of objects and their stable poses and uniform continuous distributions over the object planar pose and camera poses in a bounded region of the workspace.
- **Grasp Candidates:** $p(\mathbf{u}|\mathbf{x})$**:** A uniform random distribution over contact points on the object surface.
- **Grasp Rewards** $p(R \mid \mathbf{u}, \mathbf{x})$**:** A stochastic model of wrench resistance for the gravity wrench that is sampled by perturbing the gripper pose according to a Gaussian distribution and evaluating the contact model described in Section IV.
- **Observations** $p(\mathbf{y} \mid \mathbf{x})$**:** A depth sensor noise model with multiplicative and Gaussian process pixel noise.

Fig. 3 illustrates a subset of the Dex-Net 3.0 object and grasp dataset. The parameters of the sampling distributions and compliant suction contact model $(n, r, h, E, V, \mu, \kappa, \epsilon)$ (see Section IV) were set by maximizing average precision of the Q values using grid search for a set of grasps attempted on an ABB YuMi robot on a set of known 3D printed objects (see Section VII-A).

Our pipeline for generating training tuples is illustrated in Fig. 4. We first sample state by selecting an object at random from a database of 3D CAD models and sampling a friction coefficient, planar object pose, and camera pose relative to the worksurface. We generate a set of grasp candidates for the object by sampling points and normals uniformly at random from the surface of the object mesh. We then set the binary reward label $R = 1$ if a seal is formed and robust wrench resistance (described in Section IV-C) is above a threshold value ϵ. Finally, we sample a point cloud of the scene using rendering and a model of image noise [21]. The grasp success labels are associated with pixel locations in images through perspective projection [9]. A graphical model for the sampling process and additional details on the distributions can be found in the supplemental file.

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 [20] except that we modify the pose input stream to include the angle between the approach direction and the table normal. 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 [11]. 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 image-wise 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.

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 (CEM) [19], [20], [27]. CEM samples a set of initial candidate grasps uniformly at random from the set of surface

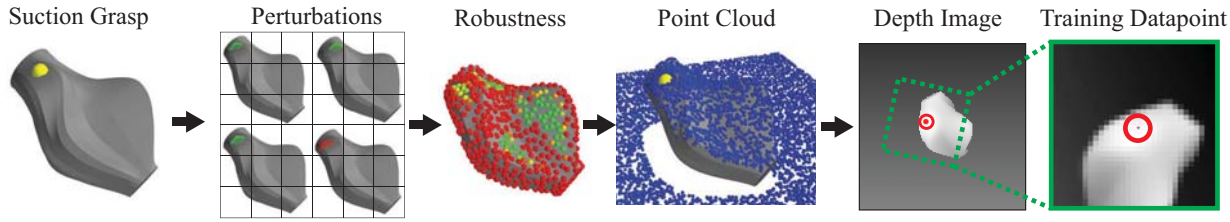

Fig. 4: Pipeline for generating the Dex-Net 3.0 dataset (left to right). We first sample a candidate suction grasp from the object surface and evaluate the ability to form a seal and resist gravity over perturbations in object pose, gripper pose, and friction. The samples are used to estimate the probability 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.

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. See the supplemental file for example grasps planned by the policy.

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.

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. 5:

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 $15mm$ diameter silicone single-bellow suction cup and a VM5-NC VacMotion vacuum generator with a payload of approximately $0.9kg$. The experimental workspace is illustrated in the left panel of Fig. 5. 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. 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 from 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 (right panel of Fig. 5). The 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, 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).** Our model without perturbations.
4) **Robust Wrench Resistance (RWR).** Our model.

The RWR metric performed best with 99% AP compared to 93% AP for WR, 89% AP for SS, and 88% for PC3D.

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

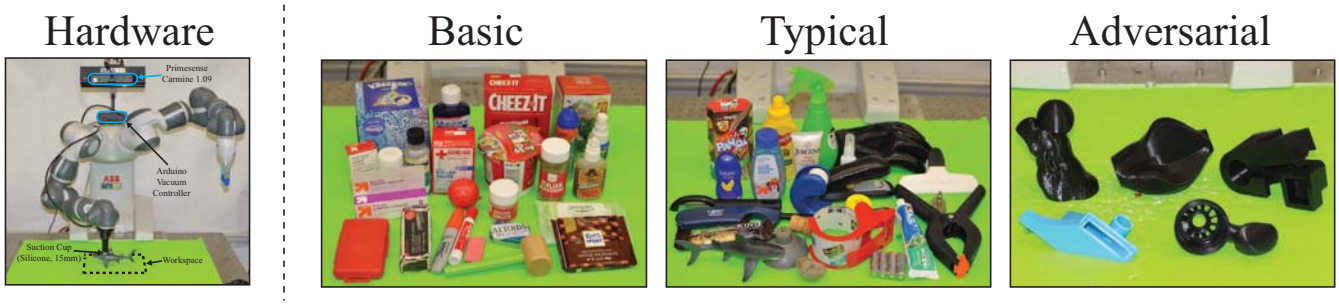

Fig. 5: **(Left)** The experimental setup with an ABB YuMi equipped with a suction gripper. **(Right)** The 55 objects used to evaluate suction grasping performance. The objects are divided into three categories to characterize performance: Basic (e.g. prismatic objects), Typical, and Adversarial.

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 (PC).** The inverse distance to the centroid for planar patches on the 3D object surface.
4) **GQ-CNN (ADV).** Our 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).** Our GQ-CNN trained on synthetic data from 3DNet [33], KIT [15], and the Adversarial objects.

Table I details performance on the Basic, Typical, and Adversarial objects. On the Basic and Typical objects, we see that the Dex-Net 3.0 policy is comparable to PC in terms of success rate and has near-perfect AP, suggesting that failed grasps often have low robustness and can therefore be detected. On the adversarial objects, GQ-CNN (ADV) significantly outperforms GQ-CNN (DN3) and PC, suggesting that this method can be used to successfully grasp objects with complex surface geometry as long as the training dataset closely matches the objects seen at runtime. The DN3 policy took an average of 3.0 seconds per grasp.

E. Failure Modes

The most common failure mode was attempting to form a seal on surfaces with surface geometry that prevent seal formation. This is partially 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 when the geometry is known precisely. In contrast, the planarity-centroid metric performs poorly on objects with non-planar surfaces near the object centroid.

VIII. Future Work

In future work we will study sensitivity to (1) the distribution of 3D object models using in the training dataset, (2) noise and resolution in the depth sensor, and (3) variations in vacuum suction hardware (e.g. cup shape, hardness of cup material). We will also extend this model to learning suction grasping policies for bin-picking with heaps of parts and to composite policies that combine suction grasping with parallel-jaw grasping by a two-armed robot. We are also working with colleagues in the robot grasping community to propose shareable benchmarks and protocols that specify experimental objects and conditions with industry-relevant metrics such as Mean Picks Per Hour (MPPH), see http://goo.gl/6M5rfw.

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 donations from Siemens, Google, Honda, Intel, Comcast, Cisco, Autodesk, Amazon Robotics, Toyota Research Institute, ABB, Samsung, Knapp, and Loccioni, Inc and by the Scalable Collaborative Human-Robot Learning (SCHooL) Project, NSF National Robotics Initiative Award 1734633. 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, Bernhard Guetl, Menglong Guo, Michael Laskey, Andrew Lee, Pusong Li, Jacky Liang, Sanjay Krishnan, Fritz Kuttler, Stephen McKinley, Juan Aparicio Ojea, Michael Peinhopf, Peter Puchwein, Alberto Rodriguez, Daniel Seita, Vishal Satish, and Shankar Sastry.

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 synthesisa 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.

	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 point-cloud-based grasping policies for 125 trials each on the Basic and Typical datasets and 100 trials each on the Adversarial dataset. We see that the GQ-CNN trained on Dex-Net 3.0 has the highest Average Precision (AP) (area under the precision-recall curve) on the Basic and Typical objects, suggesting that the robustness score from the GQ-CNN could be used to anticipate grasp failures and select alternative actions (e.g. probing objects) in the context of a larger system. Also, a GQ-CNN trained on the Adversarial dataset outperforms all methods on the Adversarial objects, suggesting that the performance of our model is improved when the true object models are used for training.

[5] 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.

[6] 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.

[7] C. Ferrari and J. Canny, "Planning optimal grasps," in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA)*, 1992, pp. 2290–2295.

[8] 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.

[9] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[10] 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.

[11] M. Jaderberg, K. Simonyan, A. Zisserman, *et al.*, "Spatial transformer networks," in *Advances in Neural Information Processing Systems*, 2015, pp. 2017–2025.

[12] 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.

[13] I. Kao, K. Lynch, and J. W. Burdick, "Contact modeling and manipulation," in *Springer Handbook of Robotics*. Springer, 2008, pp. 647–669.

[14] D. Kappler, J. Bohg, and S. Schaal, "Leveraging big data for grasp planning," in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA)*, 2015.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] J. Mahler, J. Liang, S. Niyaz, M. Laskey, R. Doan, X. Liu, J. A. Ojea, 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.

[21] 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.

[22] G. Mantriota, "Theoretical model of the grasp with vacuum gripper," *Mechanism and machine theory*, vol. 42, no. 1, pp. 2–17, 2007.

[23] D. Morrison, A. Tow, M. McTaggart, R. Smith, N. Kelly-Boxall, S. Wade-McCue, J. Erskine, R. Grinover, A. Gurman, T. Hunn, *et al.*, "Cartman: The low-cost cartesian manipulator that won the amazon robotics challenge," *arXiv preprint arXiv:1709.06283*, 2017.

[24] R. M. Murray, Z. Li, and S. S. Sastry, *A mathematical introduction to robotic manipulation*. CRC press, 1994.

[25] 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.

[26] X. Provot *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.

[27] R. Y. Rubinstein, A. Ridder, and R. Vaisman, *Fast sequential Monte Carlo methods for counting and optimization*. John Wiley & Sons, 2013.

[28] 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.

[29] 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.

[30] 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.

[31] 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.

[32] 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.

[33] 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.

[34] 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.

[35] A. Zeng, S. Song, K.-T. Yu, E. Donlon, F. R. Hogan, M. Bauza, D. Ma, O. Taylor, M. Liu, E. Romo, *et al.*, "Robotic pick-and-place of novel objects in clutter with multi-affordance grasping and cross-domain image matching," *arXiv preprint arXiv:1710.01330*, 2017.