Supplementary File for "Dex-Net 3.0: Computing Robust Robot 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]

I. Overview

This document contains additional experiments, training details for the GQ-CNN, details on computing seal formation, proofs for the suction ring contact model, and details on the stochastic environment model for dataset generation for the ICRA 2018 paper 'Dex-Net 3.0: Computing Robust Robot Suction Grasp Targets in Point Clouds using a New Analytic Model and Deep Learning".

II. 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. Performance Metrics

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

To assess performance of our robustness metric independent 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 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 results are detailed in Table I and the Success vs Attempt Rate curve is plotted in Fig. 1. A policy based on the robust wrench resistance metric achieved nearly 100%

[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

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

TABLE I: 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.

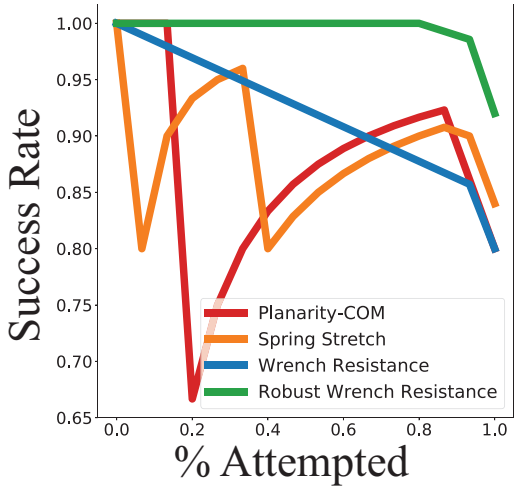

Fig. 1: Success rate vs Attempt Rate for grasp quality metrics on known 3D objects in known poses. The data was collected across 75 trials per policy on the Adversarial object dataset. The robust wrench resistance metric based on our compliant suction contact model had a 100% success rate for a large 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.

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

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. Grasp candidates were then sampled from the depth image and the grasping policy executed the most robust candidate 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 used 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 sufficiently planar patches on the 3D object surface.
4) **GQ-CNN (ADV).** A GQ-CNN trained on synthetic data from only 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 the 3DNet [14], KIT [6], and Adversarial object datasets.

Table II details performance on the Basic, Typical, and Adversarial objects, and Fig. 2 illustrates the Success-Attempt Rate tradeoff. 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. 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. The Dex-Net 3.0 policy took an average of approximately 3 seconds to plan each grasp.

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 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

	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 II: 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.

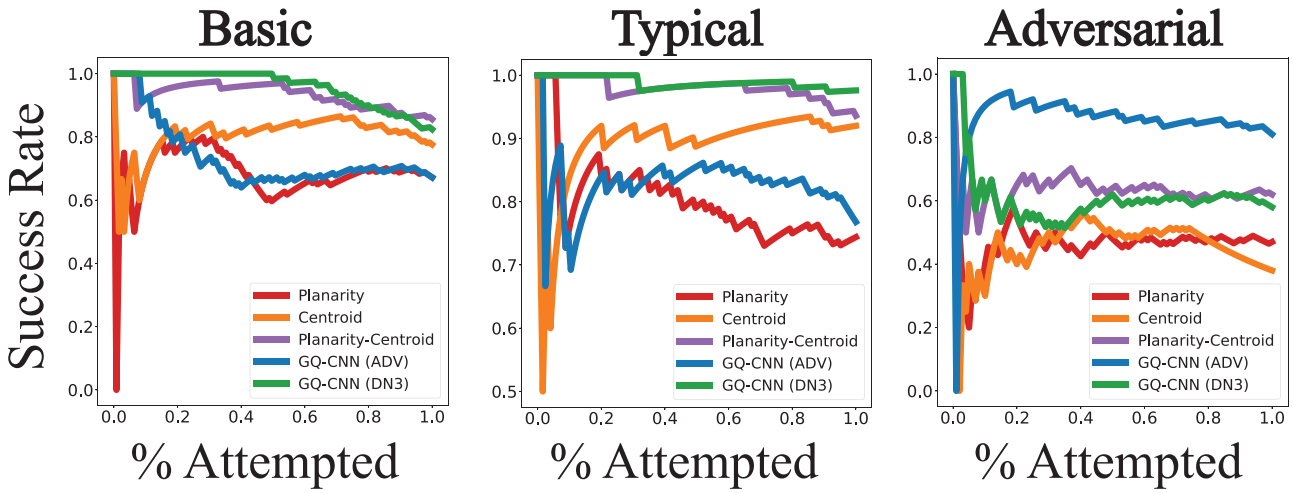

Fig. 2: Success vs Attempt Rate for 125 trials on each of the Basic and Typical object datasets and 100 trials each on the Adversarial object dataset. The GQ-CNN trained on Dex-Net 3.0 has near 100% precision on the Basic and Typical classes for a significant portion of attempts, suggesting that the 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.

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.

details the performance of each metric and the Precision-Recall curve is plotted in Fig. 3. 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:

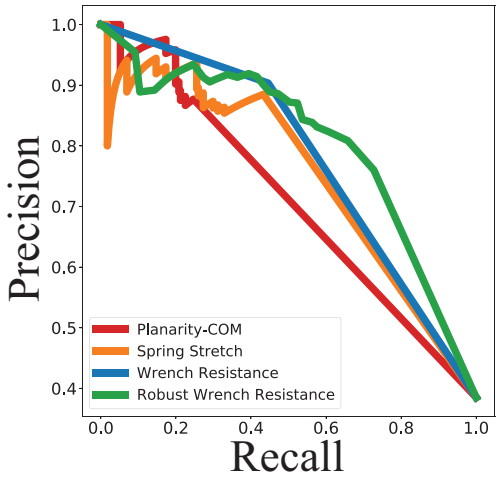

Fig. 3: Precision-Recall curve for classifying successful object lifts and transports using various metrics of grasp quality based on 3D object meshes.

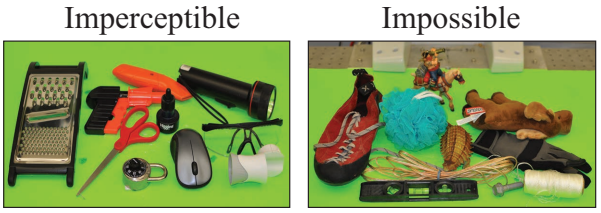

Fig. 4: Two categories that cannot be handled by any of the point-cloud based suction grasping policies. **(Left)** Imperceptible objects, which cannot be handled by the system due to small surface variations that cannot be detected by the low-resolution depth sensor but do prevent seal formation. **(Left)** Impossible objects, which cannot be handled by the system due to non-porosity or lack of an available surface to form a seal.

1) **Imperceptible Objects:** Objects with (a) surface variations less than the spatial resolution of our Primesense Carmine 1.09 depth camera or (b) specularities or 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. 4.

III. 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 complete 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. $\mathbf{v}^T(\mathbf{a}-\mathbf{p}) \leqslant 0$).
- $\mathbf{a}$ 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 [11].

IV. Suction Contact Model

The basis of contact wrenches for the suction ring model is illustrated in Fig. 5. 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 [10]. 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 [10].

Definition 1: A grasp $\mathbf{u}$ achieves *wrench resistance* with respect to $\mathbf{w}$ if $-\mathbf{w} \in \Lambda$.

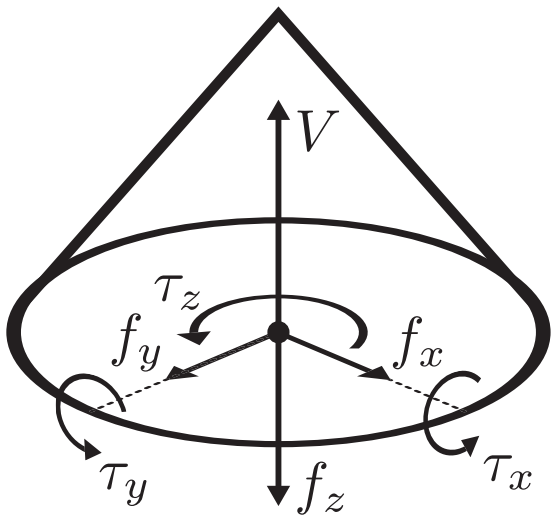

Fig. 5: Wrench basis for the compliant suction ring contact model. The contact exerts a constant pulling force on the object of magnitude V and additionally can push or pull the object along the contact z axis with force f_z. The suction cup material exerts a normal force $f_N = f_z + V$ on the object 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.

Proposition 1: Let G be the grasp map for a grasp $\mathbf{u}$. Furthermore, let $\epsilon^* = \text{argmin}_{\alpha \in \mathcal{F}} \|G\alpha + \mathbf{w}\|_2^2$. 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 $\min_{\alpha \in \mathcal{F}} \|G\alpha + \mathbf{w}\|_2^2$, 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 [5], (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* [5]. 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:

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

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:

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

We can approximate the friction limit surface by the ellipsoid [4], [5]:

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

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:

$$|f_x| \leqslant \frac{\sqrt{3}}{3}\mu f_N$$
$$|f_y| \leqslant \frac{\sqrt{3}}{3}\mu f_N$$
$$|\tau_z| \leqslant \frac{\sqrt{3}}{3} r\mu f_N$$

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:

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

where κ is the *elastic limit* or *yield strength* of the suction cup material, defined as the stress at which the material begins to deform plastically instead of linearly.

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:

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

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:

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

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 [7], [9], [12], [13], [15] 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:

$$W = \begin{bmatrix} \mathbf{e}_x & \mathbf{e}_y & \mathbf{e}_z & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{e}_z \end{bmatrix}$$
$\alpha = (f_x, f_y, f_z, \tau_z) \in \mathcal{F}$ if and only if:
$$\sqrt{f_x^2 + f_y^2} \leqslant \mu f_z$$
$$|\tau_z| \leqslant \gamma |f_z|$$

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 to 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 transport objects to a receptacle in applications such as the Amazon Picking Challenge [1], [3].

V. 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. 6 shows the precision-recall curve for the GQ-CNN validation set and the optimized 64 Conv1_1 filters, each of which is 7×7.

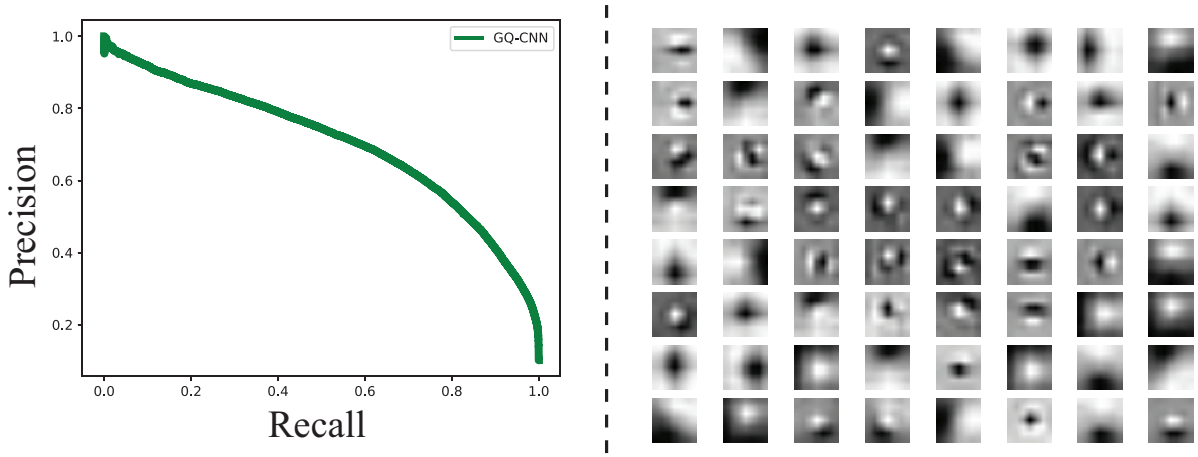

Fig. 6: **(Left)** Precision-recall curve for the GQ-CNN trained on Dex-Net 3.0 on the validation set of 552,000 pairs of grasps and images. **(Right)** The 64 Conv1_1 filters of the GQ-CNN. Each is 7×7. We see that the network learns circular filters which may be used to assess the surface curvature about the ring of contact between the suction cup and object.

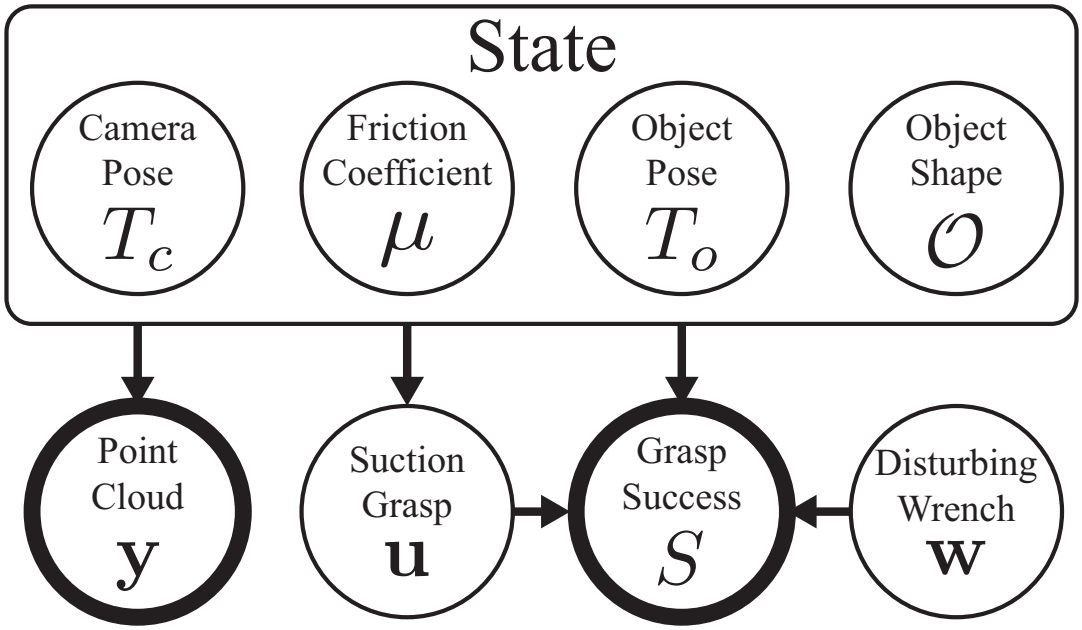

Fig. 7: A probabilistic graphical model of the relationship between the ability to resist external wrenches e.g. due to gravity under perturbations in object pose, gripper pose, camera pose, and friction.

VI. 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** $p(S \mid \mathbf{u}, \mathbf{x})$**:** A stochastic model of wrench resistance for the gravity wrench.
- **Observations** $p(\mathbf{y} \mid \mathbf{x})$**:** A sensor noise model.

Our graphical model is illustrated in Fig. 7.

Distribution	Description
$p(\mu)$	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 2.0 graphical model for generating the Dex-Net training dataset.

A. Details of Distributions

We follow the state model of [8], which we repeat here for convenience. The parameters of the sampling distributions 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 II-B).

We model the state distribution as $p(\mathbf{x}) = p(\mu)p(\mathcal{O})p(T_o|\mathcal{O})p(T_c)$. We model $p(\mu)$ 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 is $p(T_s|\mathcal{O})$ 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. [2]. 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. The parameters of the sampling distributions 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 II-B).

Our grasp candidate model $p(\mathbf{u} \mid \mathbf{x})$ 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 [8], which we repeat here for convenience. Our observation model $p(\mathbf{y} \mid \mathbf{x})$ 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 $p(S \mid \mathbf{u}, \mathbf{x})$ 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 $p(S \mid \mathbf{u}, \mathbf{x}) = p(S \mid \hat{\mathbf{u}}, \hat{\mathbf{x}}, \mathbf{w})p(\hat{\mathbf{x}} \mid \mathbf{x})p(\hat{\mathbf{u}} \mid \mathbf{u})p(\mathbf{w})$. 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 $p(\hat{\mathbf{u}} \mid \mathbf{u})$ 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 $p(\hat{\mathbf{x}} \mid \mathbf{x})$ 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 noise $\mathcal{N}(\mathbf{0}_3, 0.0025\mathbf{I}_3)$. We model $p(S \mid \hat{\mathbf{u}}, \hat{\mathbf{x}}, \mathbf{w})$ 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 $p(S \mid \mathbf{u}, \mathbf{x})$ 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] 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.

[2] 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.

[3] 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.

[4] 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.

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

[6] 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.

[7] 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.

[8] 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.

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

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

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.