Dex-Net 2.0: Deep Learning to Plan Robust Grasps withSynthetic Point Clouds and Analytic Grasp Metrics

10-20%	☑
20-40%	☑
40-60%	☑
60-80%	☑
80-100%	□
Keywork	
URL	https://arxiv.org/pdf/1703.09312.pdf
備註	
論文性質	Dex-net

I. INTRODUCTION

An alternative approach is to plan grasps using physics-based analyses such as caging [47], grasp wrench space (GWS) analysis [45], robust GWS analysis [56], or sim-ulation [26], which can be rapidly computed using Cloud Computing [28]. However, these methods assume a separate perception system that estimates properties such as object shape or pose either perfectly [45] or **according to known Gaussian distributions [35]. This is prone to errors [2], may not generalize well to new objects,** and can be slow to match point clouds to known models during execution [14].

仰賴已知的高斯分佈，無法很好的泛化

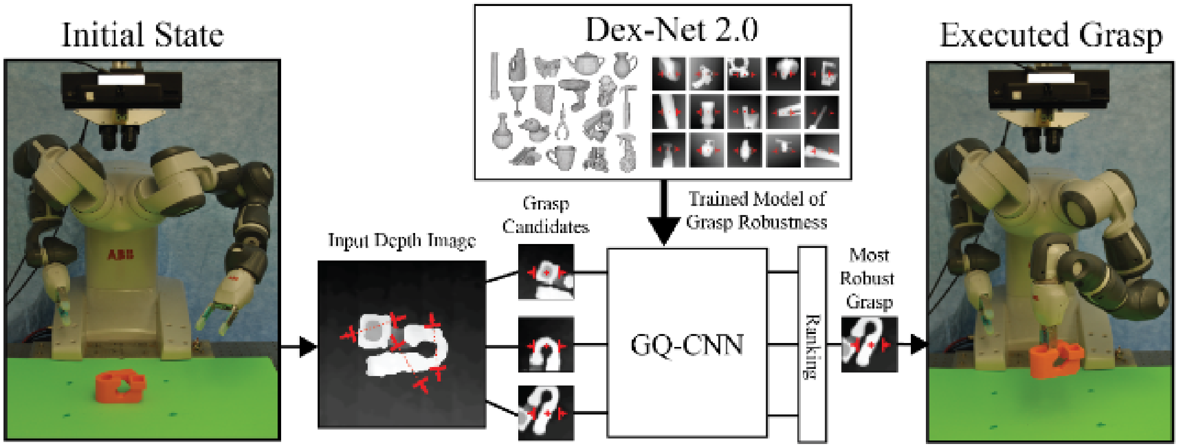


Fig. 1: Dex-Net 2.0 Architecture. (Center) The Grasp Quality Convolutional Neural Network (GQ-CNN) is trained offline to predict the robustness candidate grasps from depth images using a dataset of 6.7 million synthetic point clouds, grasps, and associated robust grasp metrics computed with Dex-Net 1.0. (Left) When an object is presented to the robot, a depth camera returns a 3D point cloud, where pairs of antipodal points identify a set of several hundred grasp candidates. (Right) The GQ-CNN rapidly determines the most robust grasp candidate, which is executed with the ABB YuMi robot.

1. 輸入深度圖
2. 得到數百個夾取點
3. 輸入gqcnn 評分
4. ABB 手臂執行

Our primary contributions are: 1) the Dexterity Network (Dex-Net) 2.0, a dataset associating 6.7 million point clouds and analytic grasp quality metrics with parallel-jaw grasps planned using robust quasi-static GWS analysis on a dataset of 1,500 3D object models, 2) a Grasp Quality Convolutional Neural Network (GQ-CNN) model trained to classify robust grasps in depth images using expected epsilon quality as supervision,where each grasp is specified as a planar pose and depth relative to a camera, and 3) a grasp planning method that samples antipodal grasp candidates and ranks them with a GQ-CNN.

1.似乎是描寫有670萬的點雲，組合成1500個物件

2.根據每次夾取，都包含的平面規劃、高度與相機資訊，以次給出分數

3.會藉由2的值做排名

In over 1,000 physical trials of grasping single objects on a tabletop with an ABB YuMi robot, we compare Dex-Net 2.0 to image-based grasp heuristics, a random forest [51], an SVM [52], and a baseline that recognizes objects, registers their 3D pose [14], and indexes Dex-Net 1.0 [35] for the most robust grasp to execute. We find that the Dex-Net 2.0 grasp planner is 3 faster than the registration-based method, 93% successful on objects seen in training (the highest of learning-based methods), and is the best performing method on novel objects, achieving 99%precision on a dataset of 40 household objects despite being trained entirely on synthetic data.

實驗設定為單個物件抓取，超過1000次。比較Dex-net2基於圖像，與隨機森林、SVM，一個是Dex-net的索引功能。

結果：3倍的快速高於註冊方法，與訓練dataset 93%的準確率。在新的40件沒看過得家庭物品上，準確率為99%。不知道有沒對測試對象、次數做詳述說明。

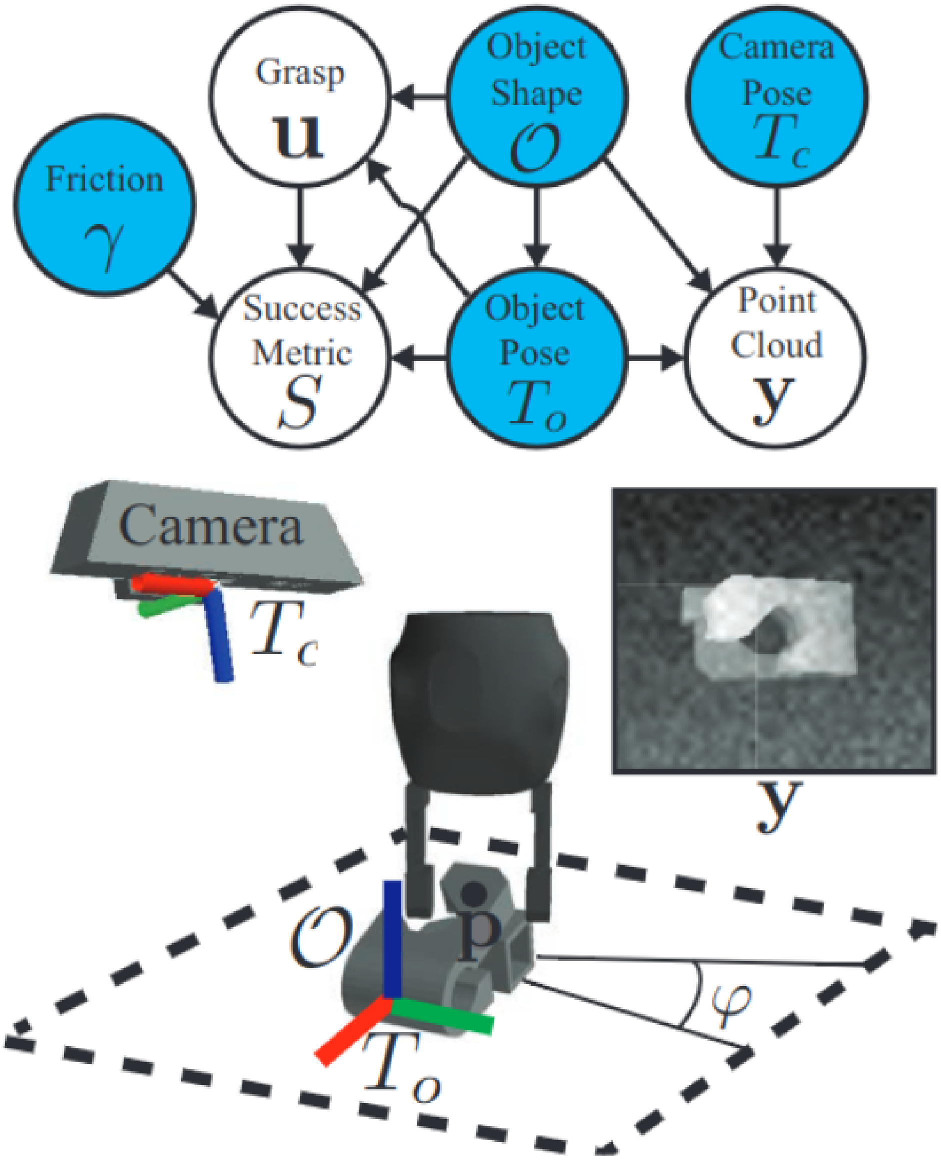


Fig. 2: Graphical model for robust parallel-jaw grasping of objects on a table surface based on point clouds. Blue nodes are variables included in the state representation. Object shapes O are uniformly distributed over a discrete set of object models and object poses T**o** are distributed over the object's stable poses and a bounded region of a planar surface. Grasps u = (p;阿法) are sampled uniformly from the object surface using antipodality constraints. Given the coefficient of friction we evaluate an analytic success metric S for a grasp on an object. A synthetic 2.5D point cloud y is generated from 3D meshes based on the

camera pose Tc, object shape, and pose and corrupted with multiplicative and Gaussian Process noise.

藍色節點會是動態，包含著狀態的代表，為輸入值。

對象O的型態不一致的散佈在一個物件的離散集。

object poses 散步在一個穩定姿態上

給予摩擦係數，我們評估一個成功指標(S)

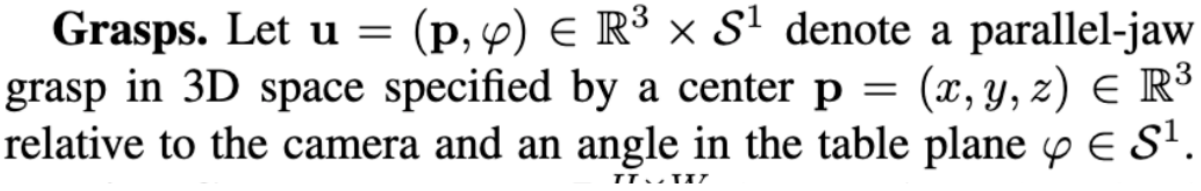
一個2.5D點雲(y)產生來自根據於相機(T)的3D網格，這還有點不明白，我以為這是相機的固定資料

multiplicative and Gaussian Process noise 損毀object shape跟pose??

B. Definitions

States.Let x= (O;To;Tc;)d

R is the coefficient of friction between the object and gripper.



Grasps. Let $\mathbf{u} = (\mathbf{p}, \varphi) \in \mathbb{R}^3 \times \mathcal{S}^1$ denote a parallel-jaw grasp in 3D space specified by a center $\mathbf{p} = (x, y, z) \in \mathbb{R}^3$ relative to the camera and an angle in the table plane $\varphi \in \mathcal{S}^1$.

p為中心點 epilson 另一個符號表示與桌面的角度

C. Objective

The estimated robustness function can be used in a grasping policy that maximizes Q over a set of candidate grasps:(y) = argmax u2C Q(u;y), where C specifies constraints on the set of available grasps, such as collisions or kinematic feasibility. **Learning Q rather than directly learning the policy allows us to enforce task-specific constraints** without having to update the learned model.

學習評估函數可以讓我嘗試最大化Q

u→ C 是規範於可約束的抓取，例如碰撞或運動學可行性

不學習policy而是Q，可以讓我們專注在特定任務

IV. LEARNING A GRASP ROBUSTNESS FUNCTION

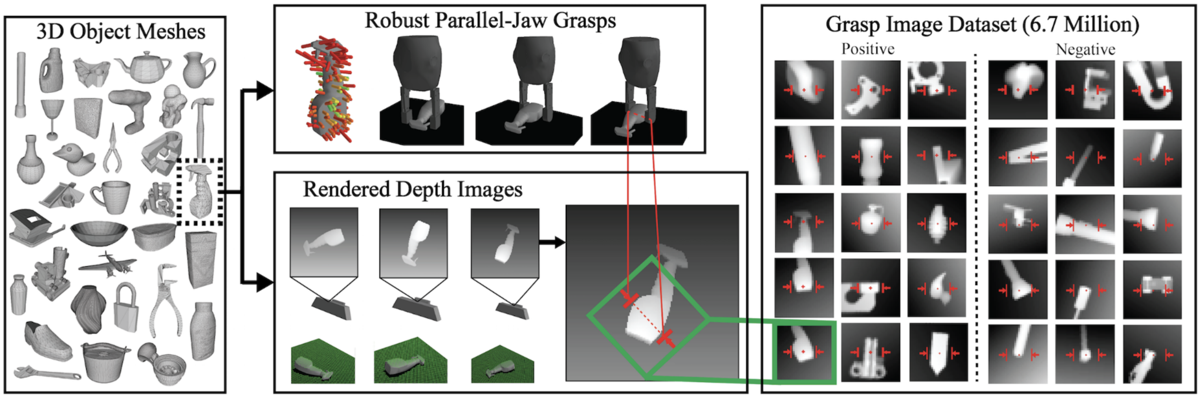


Fig.3: Dex-Net 2.0 pipeline for training dataset generation. (Left) The database contains 1,500 3D object mesh models. (Top) For each object, we sample hundreds of parallel-jaw grasps to cover the surface and evaluate robust analytic grasp metrics using sampling. For each stable pose of the object we associate a set of grasps that are perpendicular to the table and collision-free for a given gripper model. (Bottom) We also render point clouds of each object in each stable pose, with the planar **object pose and camera pose sampled uniformly at random.** Every grasp for a given stable pose is associated with a pixel location and orientation in the rendered image. (Right) Each image is rotated, translated, cropped, and scaled to align the grasp pixel location with the image center and the grasp axis with the middle row of the image, creating a 32×32 grasp image. The full dataset contains over 6.7 million grasp images.

抓取會與像素位置、取向關聯

資料庫會有協轉與裁切的32×32照片

全部的dataset包括超過 670萬的樣本

2) Database

Each mesh is aligned to a standard frame of reference using the principal axes, rescaled to fit within a gripper width of5:0cm(the opening width of an ABB YuMi gripper), and assigned a mass of 1:0kg centered in the object bounding box since some meshes are nonclosed. For each object we also compute a set of stable poses [12] and store all stable poses with probability of occurence above a threshold.

夾爪的規範： 5公分

施展直心1公斤的力

Parallel-Jaw Grasps.Each object is labeled with a set of up to 100 parallel-jaw grasps. The grasps are sampled using the rejection sampling method for antipodal point pairs developed in Dex-Net 1.0 [34] with constraints to ensure coverage of the object surface [33]. For each grasp we evaluate the expected epsilon quality EQ[42] under object pose, gripper pose, and friction coefficient uncertainty using Monte-Carlo sampling [51],

100次採樣，採樣方式為用**拒絕採樣**蒐集對立點，來自DexNet

用約束點去確保覆蓋範圍為物件表現

用蒙地卡羅採樣去評估物件位置、抓取位置、摩擦係數， 採用epsilon quality EQ

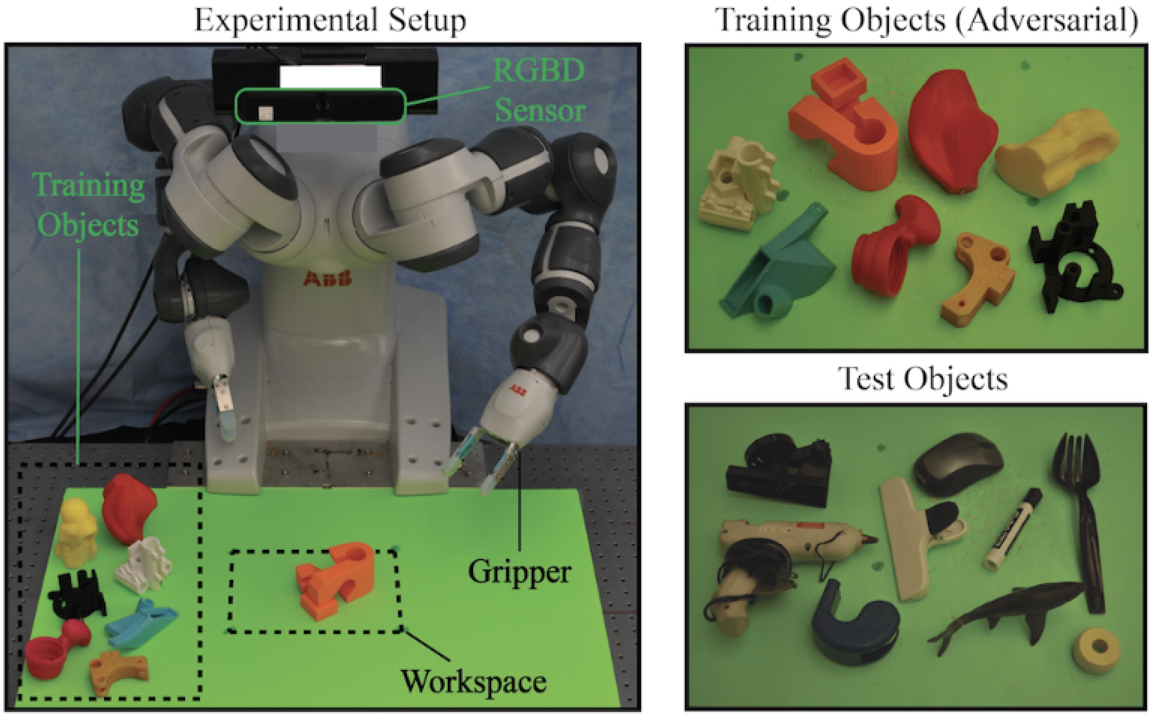


Fig.5: (Left) The experimental platform for benchmarking grasping with the ABB YuMi. We registered the camera to the robot with a chessboard before each experiment. In each trial a human operator sampled an object pose by shaking the object in a box and placing it upside down in the workspace. We then took an RGB-D image with a Primsense Carmine 1.08, filled in the image using inpainting [24], segmented the object using color background subtraction, and formed a bounding box for the detected object. The grasp planner under evaluation then planned a gripper pose and the YuMi executed the grasp. Grasps were considered successful if the gripper held the object after lifting, transporting, and shaking the object. (Top-Right) The training set of 8 objects with adversarial geometric features such as smooth curved surfaces and narrow openings for grasping known objects. (Bottom-Right) The test set of 10 household objects not seen during training.

B. Datasets

Fig. 5 illustrates the physical object datasets used in the benchmark:

1)Train:A validation set of 8 3D-printed objects with adversarial geometric features such as smooth, curved surfaces. This is used to set model parameters and to evaluate performance on known objects.

2)Test:A set of 10 household objects similar to models in Dex-Net 2.0 with various material, geometric, and spec-ular properties. This is used to evaluate generalization to unknown objects.

We chose objectsbased on geometric features under three constraints: (a) small enough to fit within the workspace, (b) weight less than 0.25kg, the payload of the YuMi, and (c) height from the table greater than 1.0cm due to a limitation of the silicone gripper fingertips.

1. 小量學習
2. 重量少於0.25
3. 因夾子關係 桌子高度高 1公分