DATA-DRIVEN GRASPING OF UNKNOWN OBJECTS

Arsalan Mousavian

CSE-571 Robotics, June 2020

HUMAN GRASPING

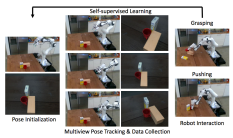
Can robots grasp as well?

Video credits: Iowa State Grocery Bagging Contest!

MODEL-BASED GRASPING

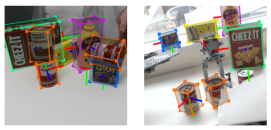
Assumes known 3D Model of Objects

- Sensing:
  - 6D Object Pose Estimation



Deng et al, ICRA 2020

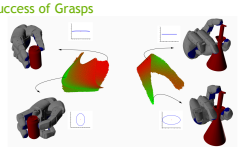
Wang et al, CVPR 2019



Tremblay et al, CoRL 2018

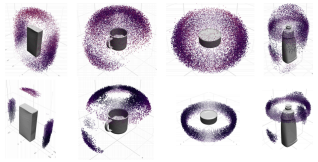
- Analyzing Success of Grasps

Force Closure



Rosales et al. RSS 2007

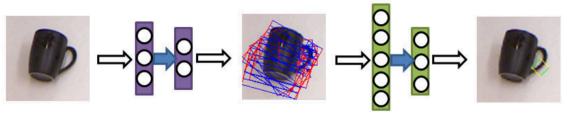
Pre-defined Grasps



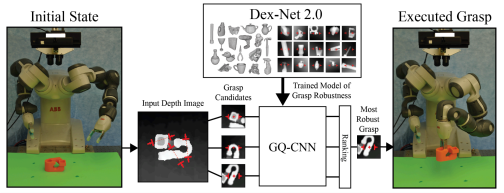
Eppner et al. ISER 2019

SUPERVISED PLANAR GRASPING

Representing grasps by oriented rectangles



Lenz et al, RSS 2013



Mahler et al, RSS 2017

RL FOR PLANAR GRASPING

Learn from large scale robot object interaction

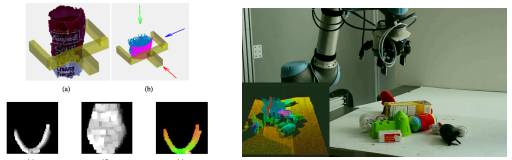
Levine et al, ISER 2016

Kalashnikov et al, CoRL 2018

ARE WE DONE?

Planar grasping is limiting.

- Limitations of planar grasping:
  - Limited workspace
  - Does not leverage the full capability of joints kinematics space.
  - Not suitable for grasping objects from enclosed spaces such as cabinets.
- 6-DoF Grasping:
  - Less constrained
  - Combinatorially larger space (6D vs 3D)

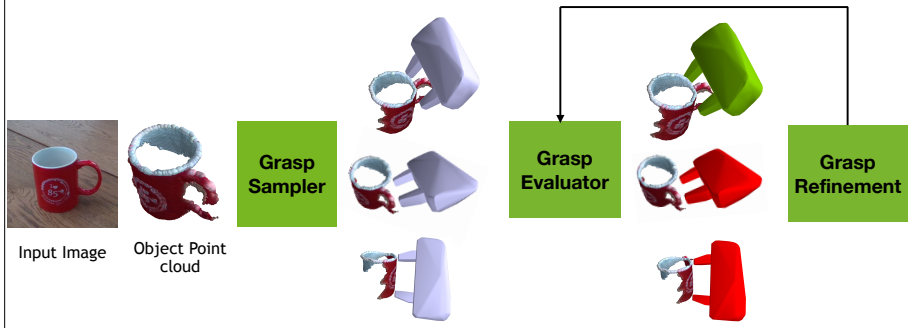


Ten Pas et al, IJRR 2017

Our Method: 6-DoF GraspNet

6-DOF GRASPNET

Generate 6D Grasp Poses from Input Point Cloud

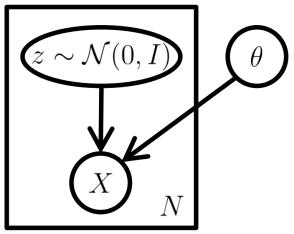


[Mousavian-Eppner-Fox, ICCV 2019]

GRASP SAMPLER

Background: Variational Auto-encoder

Objective: Having a generator model that samples from the distribution of the data: $P(X) = \int P(X \mid z; \Theta) P(z) dz$



Representation of VAE as Graphical Model
(Figure credits: Doresch, arXiv 2016)

[Kingma-Welling, ICLR 2014]

GRASP SAMPLER

Background: Variational Auto-encoder

Objective: Having a generator model that samples from the distribution of the data: $P(X) = \int P(X \mid z; \Theta) P(z) dz$

$P(X \mid z)$ is zero for most of the zs -> find likely zs with another network $Q(z \mid X)$

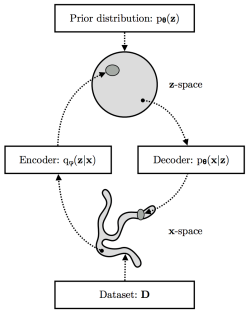


Figure credits: [Kingma-Welling, arXiv 2016]

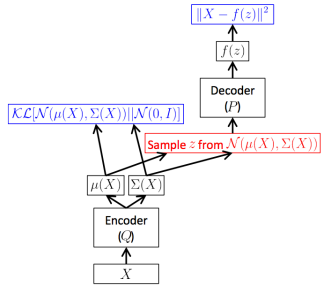


Figure credits: [Doresch, arXiv 2016]

GRASP SAMPLER

Background: Variational Auto-encoder

During inference, decoder Q is discarded and latent zs are sampled from prior distribution of z.

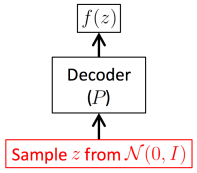
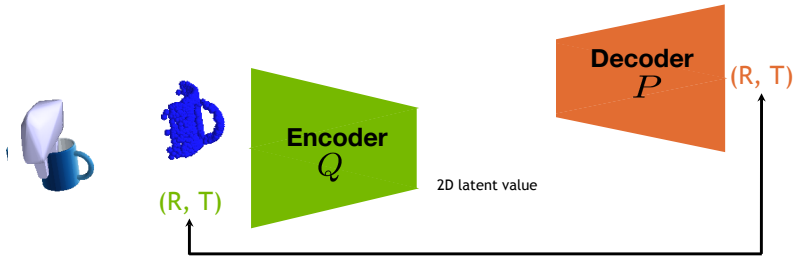


Figure credits: Doresch, arXiv 2016

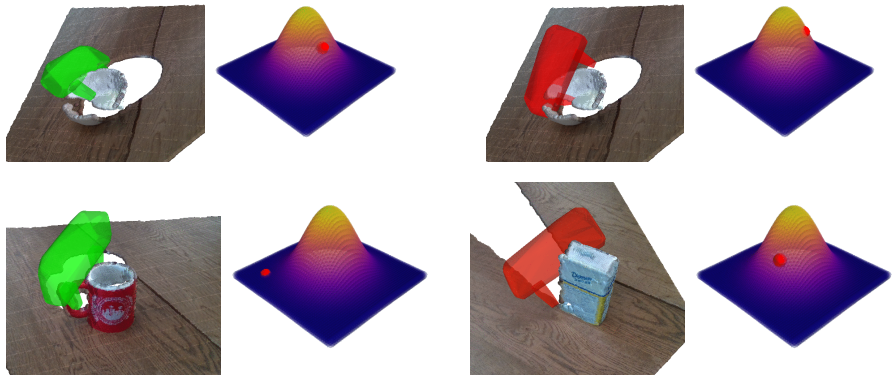
GRASP SAMPLER

Conditional VAE for Generating Grasps

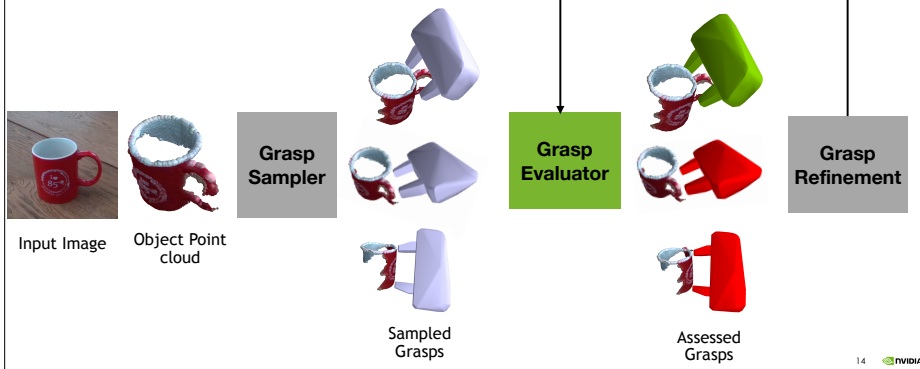


2D LATENT SPACE

Decoder generates grasps by moving through latent space



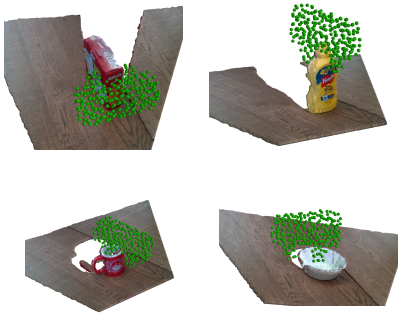
OVERVIEW



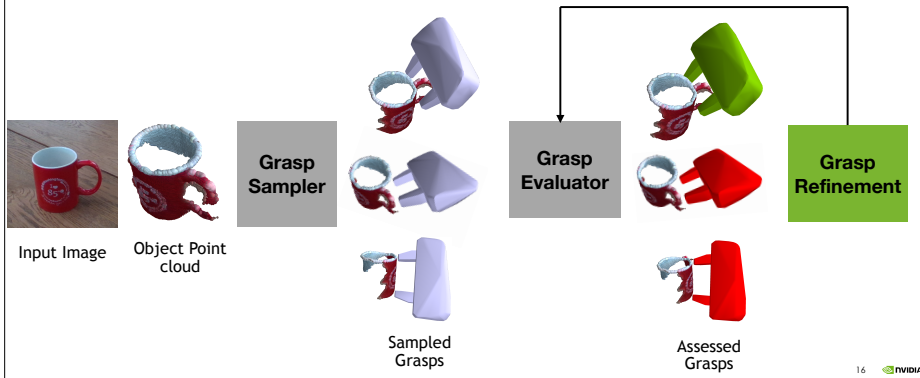
GRASP EVALUATOR

Pointnet++ model trained to discriminate successful from unsuccessful grasps

- Representation captures the relative pose of gripper and object.
- Point cloud with binary feature indicating object point or gripper point.
- Trained as binary classification to evaluate the likelihood of success for each grasp.



OVERVIEW



GRASP REFINEMENT

Evaluator provides gradient with respect to the grasp pose

TRAINING

Training is done with synthetic data

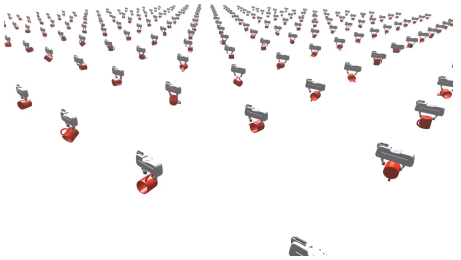
Trained on 126 random mugs, bowls, bottles, boxes, and cylinders.

Pointclouds are generated by rendering objects.

Training grasps are evaluated in NVIDIA Flex.

Tested on 17 unseen objects in real experiments.

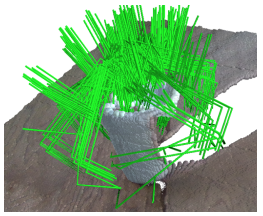
No Domain Adaptation is Needed



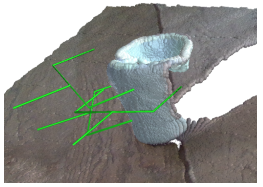
QUALITATIVE RESULTS

GENERATING DIVERSE GRASPS MATTERS

Not all predicted grasps are kinematically feasible -> Generate Diverse Grasps



6-DOF GraspNet



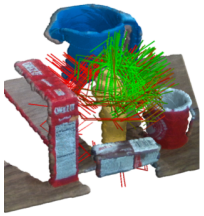
GPD [1]

	Box	Cylinder	Bowl	Mug	Average Success Rate	Success Rate
6-DOF GraspNet	**83%**	**89%**	**100%**	**86%**	**90%**	**88%**
GPD [1]	50%	78%	78%	6%	52%	47%

[1] Ten Pas et al, IJRR 2017

GRASPING OBJECTS FROM CLUTTER

APPROACH


[Murali-Mousavian-Eppner-Paxton-Fox, ICRA 2020]

APPROACH

RGB-D Observation

Single-view
RGB-D Observation

APPROACH

RGB-D Observation

Instance Segmentation

[Xie, Xiang, Mousavian, Fox, CoRL 2019]

Get Target Information with Instance Segmentation

APPROACH

3D Point Cloud

Point Cloud Observation

Assumption during learning:
Focused on collisions
between gripper and scene

APPROACH

Cropped Point Cloud

$P(G^*|X)$

grasp $g \in G$ is the SE(3) pose of the gripper
which when closing will stably lift the object

Complex G^* clutter, sensor noise, depends on
(1) Geometry of the target object
(2) Arrangement of objects in the scene

APPROACH

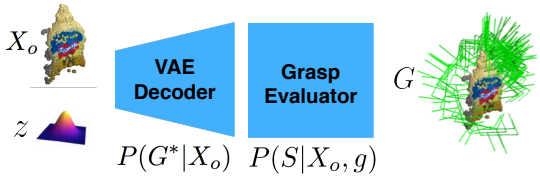
Contribution #1:
Cascaded 6-DoF Grasp Generation

(1) Object-centric grasp sampling
$P(G^*|X_o)$

(2) Clutter-centric evaluation with CollisionNet
$P(C|X, g)$

APPROACH

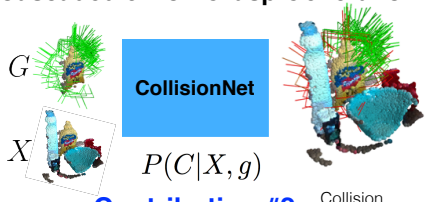
Cascaded 6-DoF Grasp Generation



(1) Object-centric grasp sampling with VAE

APPROACH

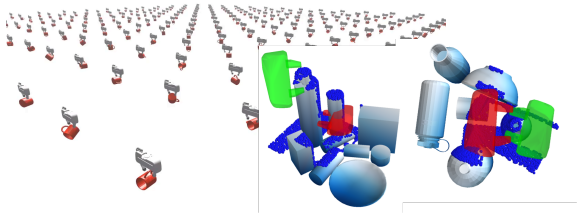
Cascaded 6-DoF Grasp Generation



Contribution #2: (2) Clutter-centric evaluation with CollisionNet, a learnt collision-checker

APPROACH

Training in Simulation

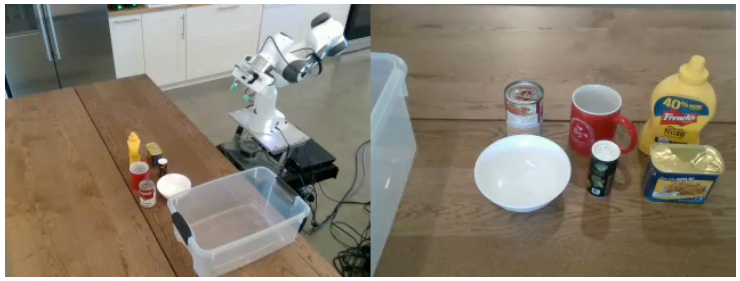


Object-centric Grasps

Collision labels and Point Clouds rendered from simulated clutter

EXPERIMENTAL EVALUATION

Real Robot Experiments



Grasp performance of **80.3%** on 23 unknown objects in clutter (for a total of 9 scenes) on a real robot, outperforms baseline by **17.6%**

CollisionNet outperforms a voxel-based approach in robot experiments (by **19.6%**)

Transfer to real robot and data!

EXPERIMENTAL EVALUATION

Application: Remove Blocking Objects

Target object specified by human user

EXPERIMENTAL EVALUATION

Application: Remove Blocking Objects

Target object is initially not reachable;

grasps will collide with surrounding clutter

EXPERIMENTAL EVALUATION

Application: Remove Blocking Objects

Blocking objects are ranked using CollisionNet

(**red** has the highest score and **green** is the lowest)

EXPERIMENTAL EVALUATION

Application: Remove Blocking Objects

New goal: remove the object with the highest blocking score

EXPERIMENTAL EVALUATION

Application: Remove Blocking Objects

Blocking object is removed from the scene

EXPERIMENTAL EVALUATION

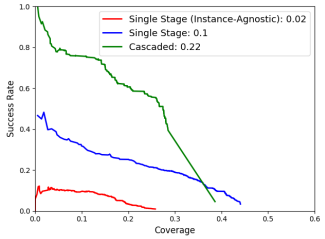
Application: Remove Blocking Objects

Target object is now reachable
Grasp success!
and can be retrieved

EXPERIMENTAL EVALUATION

Ablations in Simulation

Success Rate:
Proportion of generated grasps that lift the target object ***and*** do not collide with clutter

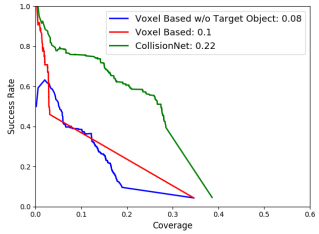


Coverage:
Proportion of ground truth grasps that are close to generated grasps

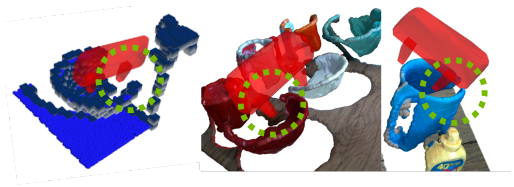
Contribution #1:
Cascaded grasp generation outperforms
1) single-stage by AUC 0.12
2) instance-agnostic approach by AUC 0.20

EXPERIMENTAL EVALUATION

Ablations in Simulation



False Positives from Voxel-based approach



Contribution #2:
CollisionNet outperforms traditional voxel-based collision checking by AUC 0.12

[Hornung et. al. Autonomous Robots 2013]

CONCLUSIONS

- New approach to generate 6-DoF grasps from object point cloud for unknown objects.
- The method does not need any semantic information about the objects -> scalable.
- Works directly on raw sensory data -> more robust.
- Limitations and Future Works:
  - Closing the loop
  - Consider Robot trajectory during grasp generation
  - Use learned modules in task planning applications

REFERENCES

- 6-DoF Grasping:
  - “6-DoF GraspNet: Variational Grasp Generation for Object Manipulation”, Mousavian et al. ICCV 2019
  - “6-Dof Grasping for Target Driven Object Manipulation”, Murali et al. ICRA 2020
- Instance Segmentation:
  - “The best of both modes: Separately leveraging RGB and Depth for Unseen Object Instance Segmentation”, Xie et al. CoRL 2019
- Variational Auto-encoder:
  - “Tutorial on Variational Autoencoders”, Doresch, arXiv 2016
  - “An introduction to Variational Autoencoders”, Kingma et al, arXiv 2019
- Neural network for point cloud:
  - “Pointnet++: Deep Hierarchical Feature Learning on Point Set in a Metric Space”, Qi et al. NeurIPS 2018