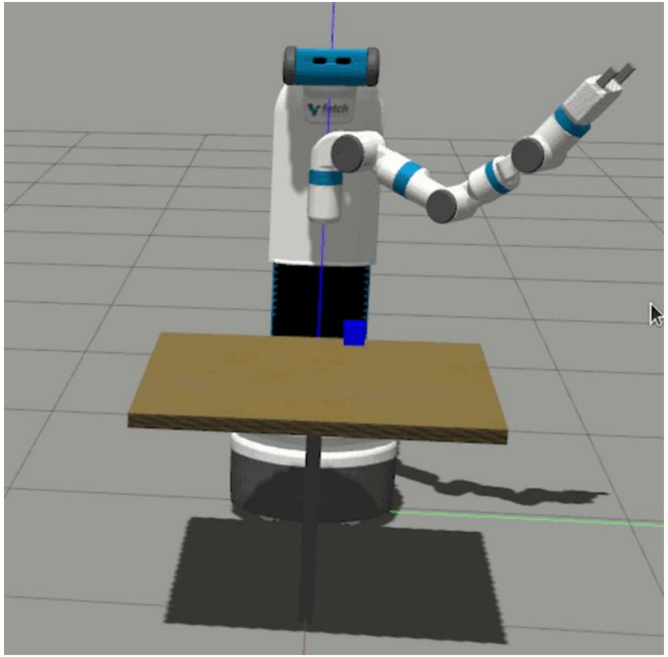
AskGrasp: Is this grasp suitable for performing a task on a particular object?

Abhishek R, Jishnu P, Harish G, Priyanshi S

Team-11

CS6301.001-Introduction to Robot Manipulation and Navigation

Till now



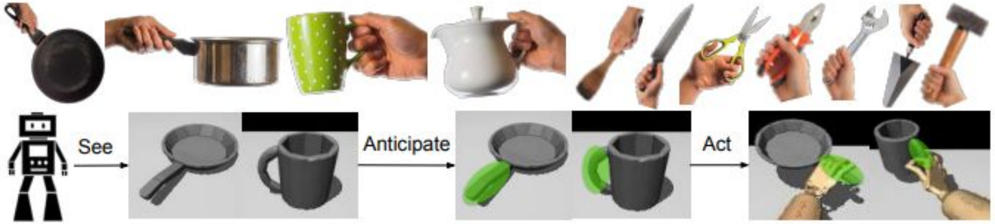
Given an object(here **cube**), a Robot picks the object.

Given an object and a set of predefined grasps, robot picks an objects by selecting a feasible grasp from the set based on the environment.

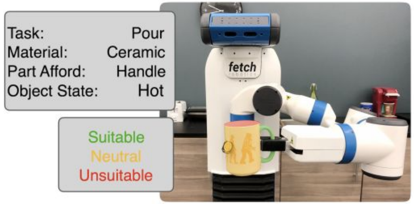
Now what?

How to make the robot capable of doing various tasks using the object?

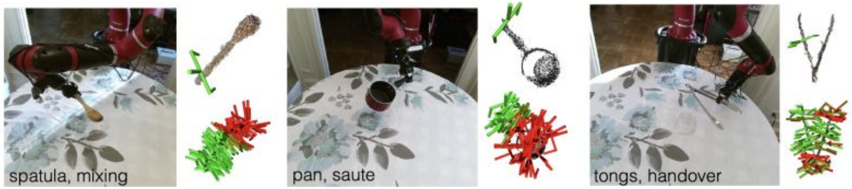
Few Approaches



(a) Object parts and Affordance detection
https://arxiv.org/pdf/2009.01439.pdf



(b) Semantic Knowledge in Robotics
https://arxiv.org/pdf/1909.11142.pdf

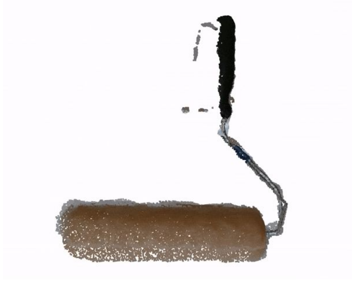


(c) Semantic knowledge: Same Object, Different Grasps
https://arxiv.org/pdf/2011.06431.pdf

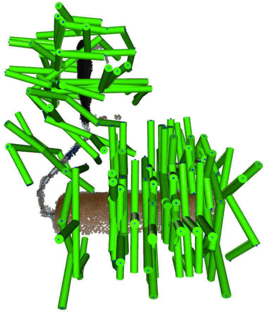
Goal: Classify the given set of Grasps on an Object based on whether it can perform the given Task

Fetch-Image: https://fetchrobotics.com/fetch-mobile-manipulator
Other media: https://github.com/adithyamurali/TaskGrasp

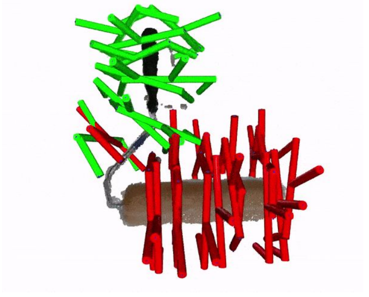
(1) Robot views the scene



(2) Get Object point cloud (here: Paint Roller)

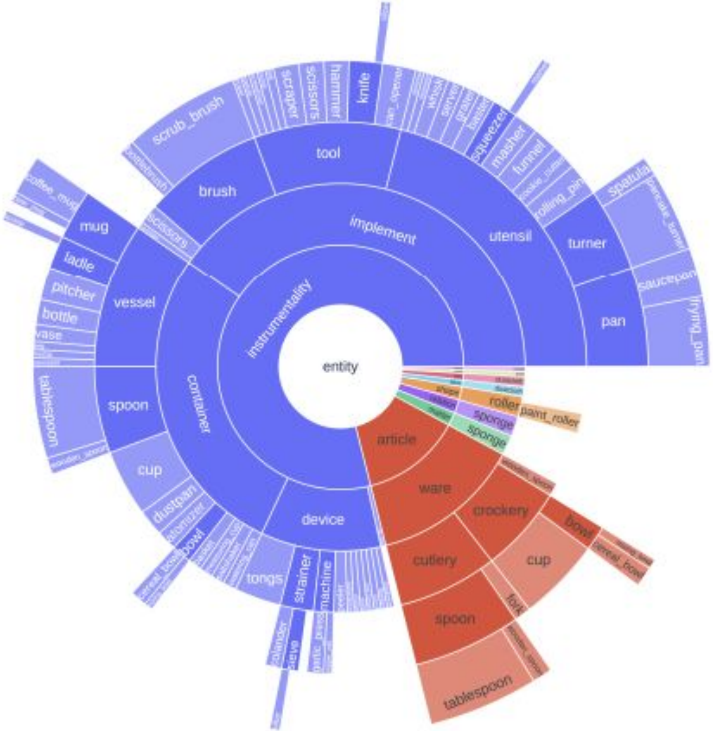


(3) Sample possible Grasps on the Object



(4) Good/Bad Grasps conditioned on Task (here: Paint)

Dataset: TaskGrasp



Source: TaskGrasp
https://arxiv.org/pdf/2011.06431.pdf

250K task-oriented grasps

56 tasks

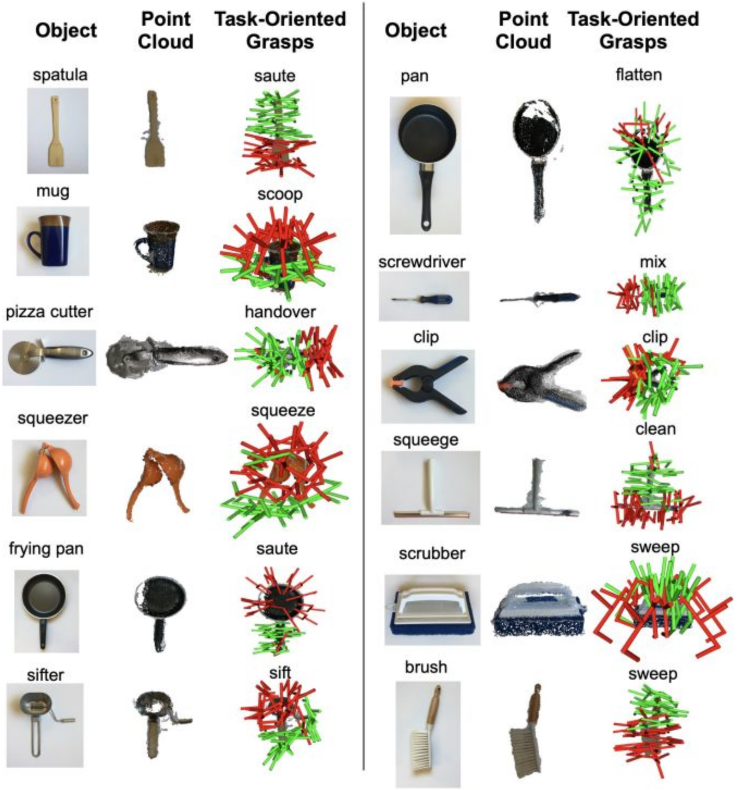
191 household and kitchen objects

- 75 distinct object categories

RGB-D information

Grasp Type - SE(3)

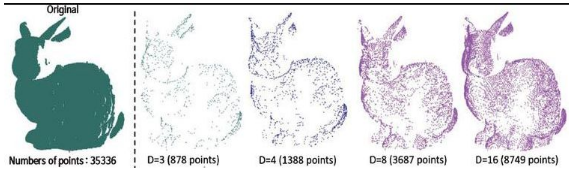
Dataset: TaskGrasp cont…



Source: TaskGrasp
https://arxiv.org/pdf/2011.06431.pdf

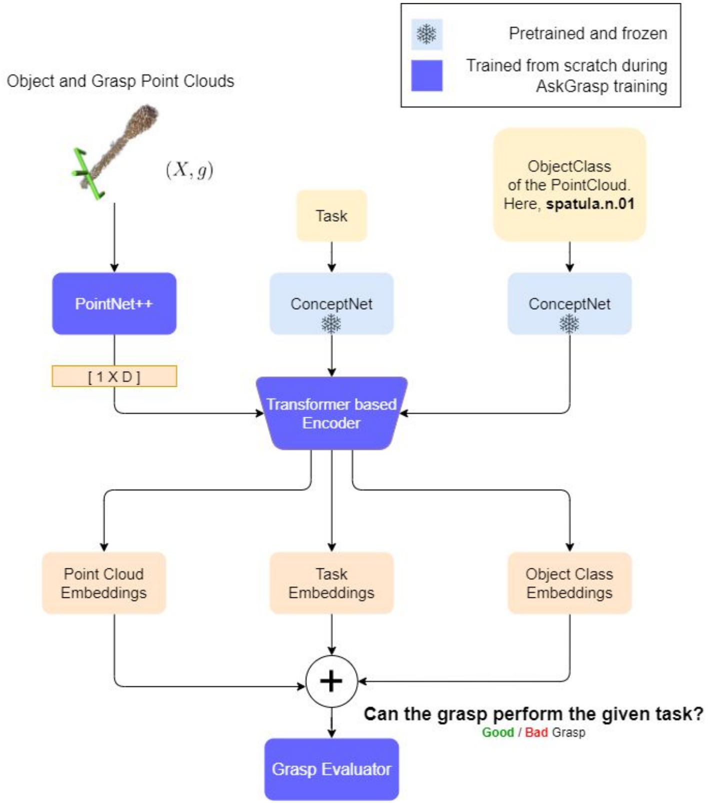
Point Cloud:

- A point cloud is a collection of tiny individual points captured using a 3D laser scanner and plotted in 3D space.
- The denser the points, the more detailed the representation.



https://content.iospress.com/articles/journal-of-intelligent-and-fuzzy-systems/ifs182742

Our AskGrasp Model



❄ Pretrained and frozen
🟦 Trained from scratch during AskGrasp training
Object and Grasp Point Clouds
(X, g)
Task
ObjectClass of the PointCloud. Here, **spatula.n.01**
PointNet++
ConceptNet
ConceptNet
[1 X D]
Transformer based Encoder
Point Cloud Embeddings
Task Embeddings
Object Class Embeddings
+
Can the grasp perform the given task?
Good / Bad Grasp
Grasp Evaluator

- **PointNet++:**
  - Encode 3D point clouds into meaningful embedding considering local and global features
- **ConceptNet:** https://arxiv.org/abs/1612.03975
  - used to create word embeddings -- representations of word meanings as vectors, similar to word2vec, GloVe, or fastText, but better.
- **Transformer based Encoder:** https://arxiv.org/abs/1706.03762
  - To leverage the self-attention mechanism
  - Useful blog: https://jalammar.github.io/illustrated-transformer
- **Grasp Evaluator:**
  - MLP with (256, 256, 256)
- **Loss**
  - Binary Cross Entropy

Model Hyperparameter used for training

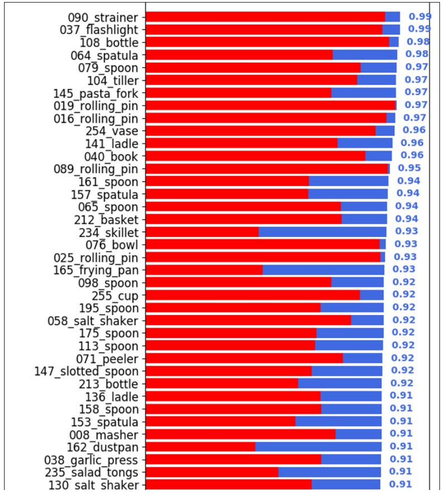
- **Batch_size** = 16
- **Num_points** = 4096
- **Epochs** = 200
- **Adam_Optimizer_lr** = 1e-4
- **D_models** = 300
- **Nhead** = 6
- **Dim_feedforward** = 1024
- **Dropout** = 0.2
- **Layers** = 4

Results

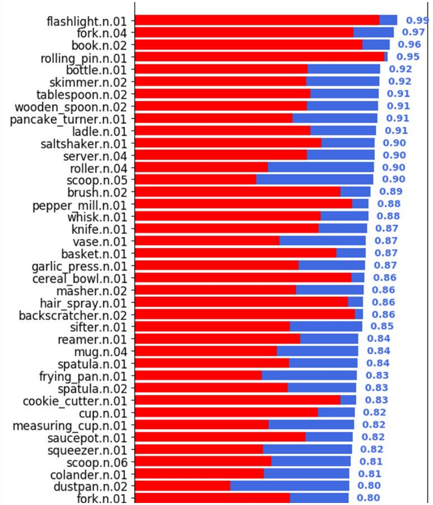
Model	**Test Performance(mAP)**		
	Instance	**Classes**	**Tasks**
Random	59.75	60.28	54.76
SGN	78.51	75.08	68.8
SGN+word embedding	79.74	77.91	**74.36**
GCNGrasp	**80.25**	**77.94**	73.71
AskGrasp (ours)	77.67	76.43	63.10

Object Instance Generalization

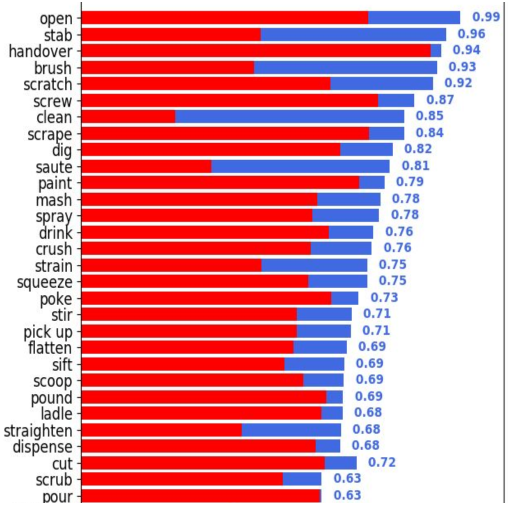
Results Continued...



Instance (mAP = 77.67%)



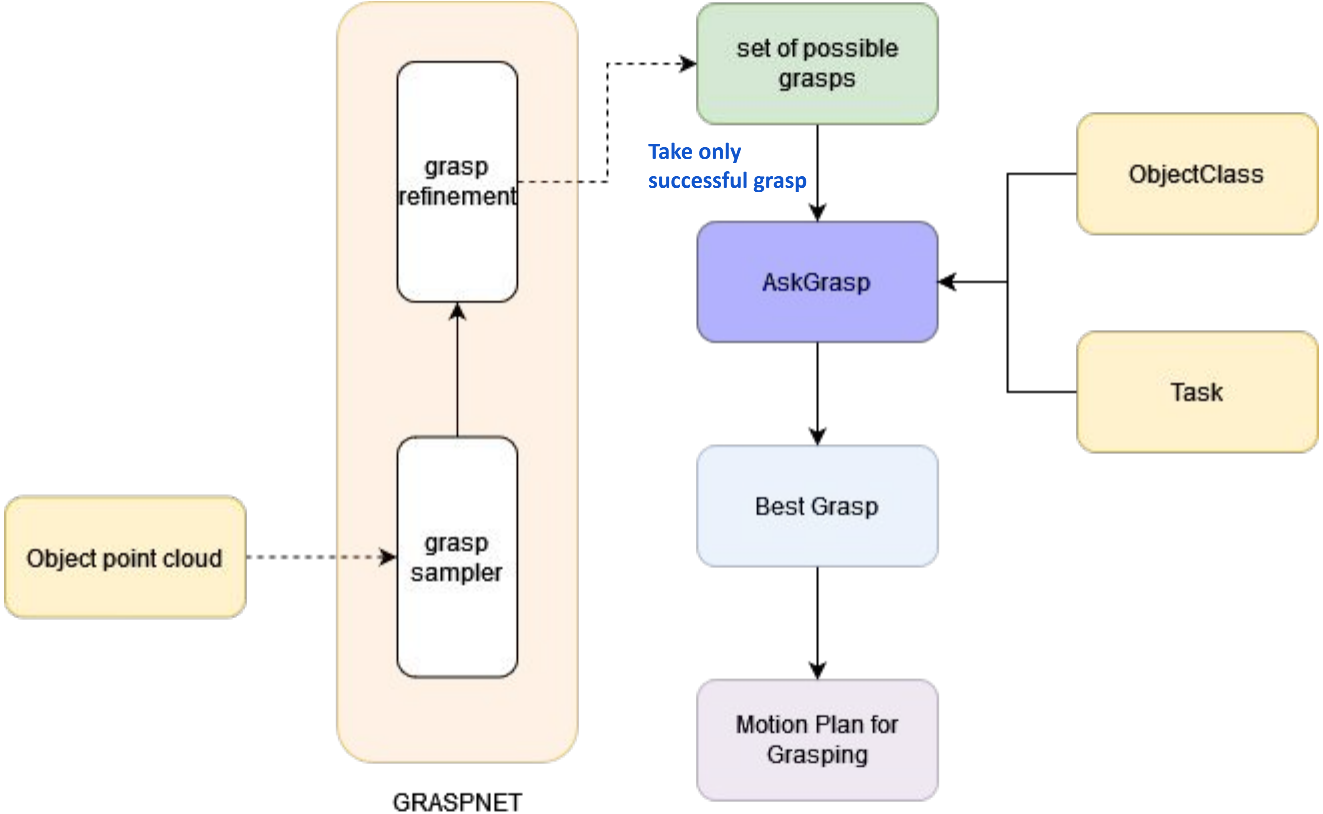
Class (mAP = 76.43%)



Task (mAP = 63.10%)

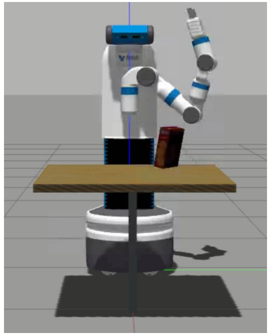
Inference

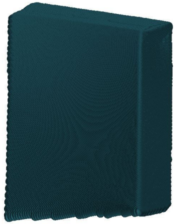
Workflow



https://arxiv.org/abs/1905.10520

Step-1: Point Cloud Generation





Open3D Point Cloud visualization

The point clouds are used in grasp generation and fed into the AskGrasp model for inference.

Step-2: Grasps Generation

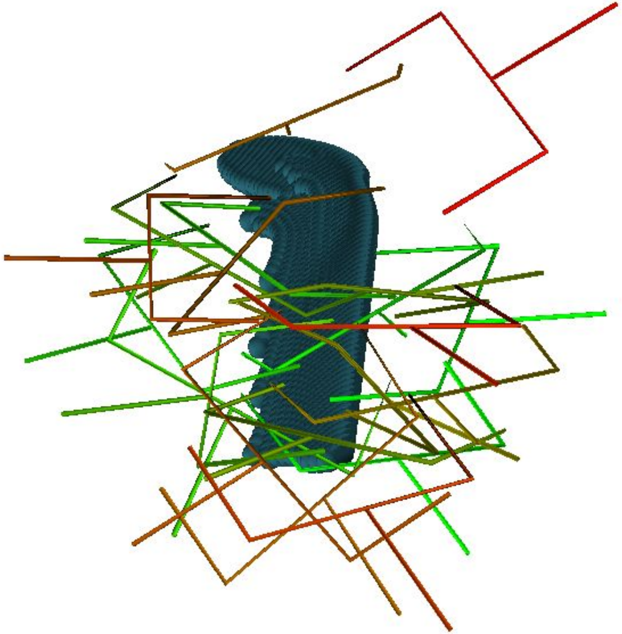
Grasp:

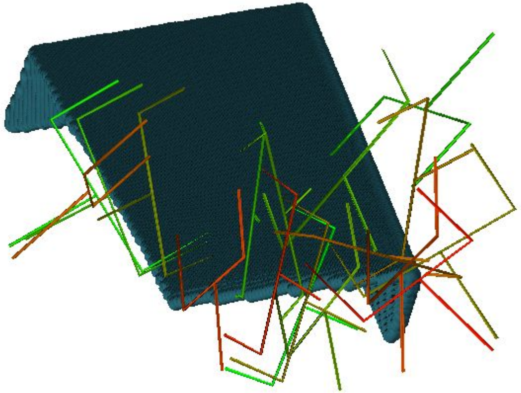
It represents a final pose of gripper link of fetch robot which will result in a successful hold over the object in camera frame.

Grasp Generation:

- Made use of 6-DOF GraspNet which produces possible grasps for a given input of point clouds generated from depth camera.
- Pytorch Implementation: https://github.com/jsll/pytorch_6dof-graspnet
- **Input** : Point cloud from depth camera of fetch robot simulation in gazebo
- **Output** : Set of possible grasps generated by combination of grasp sampler and grasp refinement network trained from GitHub

Examples for generated Grasps





Green/Red colors indicate feasible/infeasible grasps

NOTE: These grasps are not conditioned on tasks

Questions?