Deep Learning Based Grasp Pose Prediction for Hand-held Objects on Table-Top Setting

Sigournie Brock, Gonzalo Valdenebro, Christopher Porter, and Md Alimoor Reza

Drake University

Motivation

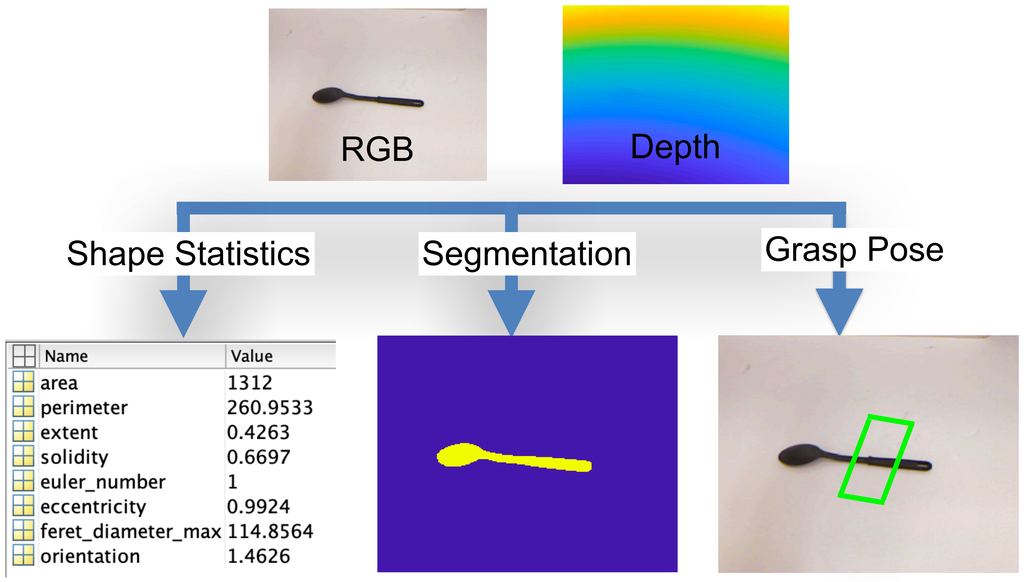
- We addressed the problem of grasp pose prediction of an object by a robotic manipulator
- We extended an existing grasp-detection model[1], which operates on a pair of RGB and depth as input images to predict the grasp pose of different objects commonly found on top of table surfaces

Problem Statement

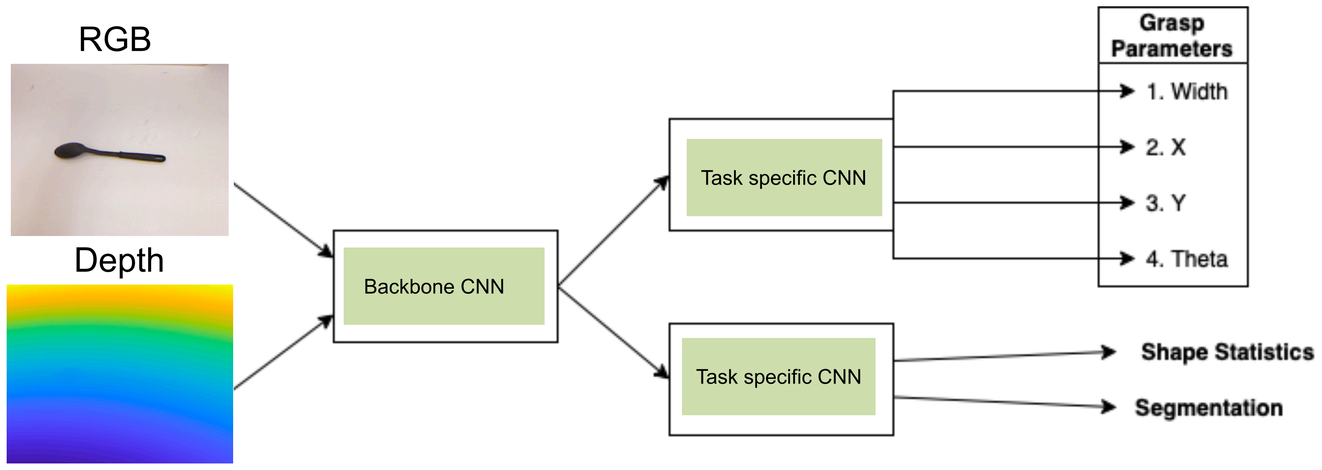
- Given an input RGB-D image, we learn a deep-learning model that predicts grasp pose of an object in terms of four parameters: (1) *width*, (2) *x-location*, (3) *y-location*, and (4) *angle*
- Unlike the work of [1], we train our model in a Multi-Task Learning (MTL) framework where new tasks were incorporated besides grasp pose prediction

Multi-Task Learning

- Tasks help each other to boost their individual performance when trained in a Multi-Task learning (MTL) framework[3,4,5]
- Two additional prediction tasks are (a) shape statistics prediction and (b) segmentation prediction
- Seven shape statistics: i) area, ii) perimeter, iii) extent, solidity, iv) solidity, v) eccentricity, vi) feret-diameter, and vii) orientation from the ground-truth segmentation mask of each object



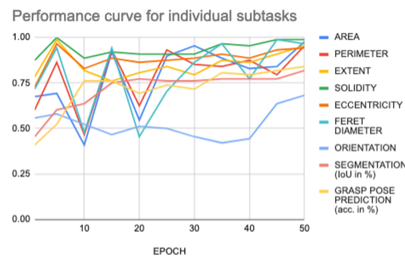
Deep Neural Network Architecture



Results

- We conducted experiments on the Cornell Dataset[2]. It contains 800 images of several different graspable objects on tabletop indoor scenes
- The model was developed in PyTorch framework

EPOCH	SHAPE STATISTICS (acc. in %)							SEGMENTATION (IoU in %)	GRASP POSE PREDICTION (acc. in %)
	AREA	PERIMETER	EXTENT	SOLIDITY	ECCENTRICITY	FERET DIAMETER	ORIENTATION		
1	0.67	0.60	0.78	0.88	0.73	0.72	0.56	0.45	0.41
5	0.69	0.86	0.99	1.00	0.97	0.94	0.58	0.60	0.52
10	0.41	0.47	0.82	0.89	0.83	0.48	0.52	0.64	0.76
15	0.92	0.93	0.76	0.92	0.89	0.94	0.47	0.75	0.76
20	0.55	0.63	0.81	0.91	0.86	0.45	0.51	0.77	0.69
25	0.90	0.93	0.84	0.91	0.88	0.70	0.50	0.76	0.74
30	0.95	0.85	0.80	0.91	0.89	0.86	0.45	0.76	0.72
35	0.89	0.84	0.88	0.97	0.91	0.97	0.42	0.77	0.81
40	0.83	0.88	0.86	0.95	0.89	0.77	0.44	0.77	0.80
45	0.84	0.80	0.91	0.99	0.93	0.99	0.64	0.77	0.82
50	0.98	0.95	0.95	0.99	0.94	0.97	0.68	0.82	0.84



Future Work

- Conduct experiments with more shape statistics in order to determine what modality works best for optimal grasp detection
- Conduct experiments with different CNN backbones (ResNet, VGG, ConvNeXt[6])
- Incorporate new loss function that learns weight factors of different tasks in our MTL framework[4]

References

1. Antipodal Robotic Grasping using Generative Residual Convolutional Neural Network - S. Kumar et al. IROS'20
2. Cornell Grasping Dataset: https://www.kaggle.com/datasets/oneoneliu/cornell-grasp
3. Multi-Task Learning with Deep Neural Networks: A Survey - M. Crawshaw (arXiv'20)
4. Multi-Task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics - A. Kendall et al. CVPR'18
5. BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning - F. Yu et al. CVPR'20
6. A ConvNet for the 2020s - arXiv'22

Acknowledgements: This work was partially supported by NASA Iowa Space Grant Consortium