

# 6.857 Final Project: Milestone 1

Laura Jarin-Lipschitz, Sebastiani Aguirre Navarro and Rachel Holladay

October 5, 2017

Reliable grasping is at the heart of most robotic manipulation tasks. The goal of our project is predict grasp success to enable reliable grasping. We imagine a scene with a robotic manipulator, a goal object to be grasped, and a camera providing a view of the goal object. We want to decide how to grasp that goal object such that our probability of a successful grasp is maximized. Generating grasps given a view of the scene is its own difficult research area [1, 2, 3]. Therefore, for the context of this machine learning-focused project, we will assume we have a black box way to generate grasps drawn from existing, open source packages, as detailed in the dataset description below. Therefore, our learning input is an image and proposed grasp, and our desired output is a metric of probability of grasp success. By learning this relation, we will score and then rank all possible grasps for an object, allowing execution of the best grasp.

The relatively recent interest in the intersection of large scale learning and manipulation has led to the publication of several recent grasping datasets. We have investigated and currently plan to use both DexNet 2.0 [4] and BigBIRD [5].

1. DexNet 2.0 provides 6.7 million sets of synthetic point clouds, grasps and analytical grasp metrics. Here, the grasps are provided by the data set, eliminating the need to generate them. The grasps are for a parallel jaw-gripper and the dataset is based off of millions of 3D models of common objects on a table.
2. BigBIRD is a dataset across 125 objects placed on a table, where each object has 600 images and 3D point clouds. This creates a dataset of 75,000 samples. While this is a smaller set than DexNet, the point clouds are not synthetic. ten Pas et al utilized this dataset to generate grasps and classify their success rate [6]. Their grasp generation method for a parallel jaw-gripper is open source, allowing us to use it as a black box.

Our hope is that we can utilize both the datasets, nearly interchangeably with respect to our architecture, to compare across synthetic and real point cloud data and to compare the effect of the size of the dataset on the quality of results. Following the methods of the cited papers, we plan to create and

apply a CNN architecture, but we will not replicate either architecture from the literature. We plan to evaluate our approach based on our performance on a hold-out set of test data. Due to time constraints, we do not expect to evaluate our prediction method on a physical robotic platform, and instead will use a model-based robotics simulator such as Gazebo.

## References

- [1] L. Pinto and A. Gupta, "Supersizing self-supervision: Learning to grasp from 50k tries and 700 robot hours," in *ICRA*, pp. 3406–3413, IEEE, 2016.
- [2] I. Lenz, H. Lee, and A. Saxena, "Deep learning for detecting robotic grasps," *IJRR*, vol. 34, no. 4-5, pp. 705–724, 2015.
- [3] Y. Jiang, S. Moseson, and A. Saxena, "Efficient grasping from rgb-d images: Learning using a new rectangle representation," in *ICRA*, pp. 3304–3311, IEEE, 2011.
- [4] 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," *arXiv preprint arXiv:1703.09312*, 2017.
- [5] A. Singh, J. Sha, K. S. Narayan, T. Achim, and P. Abbeel, "Bigbird: A large-scale 3d database of object instances," in *ICRA*, pp. 509–516, IEEE, 2014.
- [6] A. t. Pas, M. Gualtieri, K. Saenko, and R. Platt, "Grasp pose detection in point clouds," *arXiv preprint arXiv:1706.09911*, 2017.