6.857 Final Project: Milestone 3

Sebastiani Aguirre Navarro and Rachel Holladay

October 26, 2017

1 Software Infrastructure

Our goal is to use machine learning techniques to predict the probability of successfully grasping an unknown object with a robotic arm. The infrastructure of our project is broken into two sections: the data generation and the learning.

Our data will come as point clouds from the Big Bird dataset [1], which provides depth maps, RGB-D images and meshes. We have already secured access to this dataset, and while there are 125 objects in the data set, we do not expect to use all of them. The next step in the data generation process is creating labeled grasps from the point cloud. We will be using the open source grasp generation and labeling process package "Grasp Pose Generator (GPG)" from [2]. The package detect 6 degree of freedom poses for parallel jaw grippers from point clouds. The code is available through a Github repository and has a few dependencies, such as ROS and Eigen. Our goal is to use this package to generate data as soon as possible.

Given our labeled data, when can then turn to our learning algorithms. We will use Tensorflow and Keras to implement our learning algorithms. Keras specifically containes predefined many of the common layers used in deep neural networks. This will prove useful to run initial prototypes that build upon these common layers. Eventually, we will have to tweak the layers or add layers that are not predefined in Keras. This is where we will turn to Tensorflow. While Tensorflow already has many basic predefined layers, prototyping in it can be tedious. However Tensorflow provides the fundamental tools for defining custom operations and layers. There is also an existing "Model Zoo" for Keras and Tensorflow. Since many of the proposed architectures have been already trained with the purpose of detecting or classifying objects, one strategy is to properly modify the input layers and perform Transfer Learning to see if we can reduce the training time significantly in our grasper network when tuning it on the RGBD data. We will also write from scratch several prototypes that apply the concepts of ResNet, Network in Network and Inception so that we can compare it to other transfered off the shelf models.

2   Open Questions

From an infrastructure perspective, we are still discussing the best representation of the data for our learning algorithm. There is a lengthy discussion and comparison of representing grasp candidates in [2] with respect the volume that the grasp occupies. We will use their reported experiences alongside our own representation to determine the best represetation.

3   Division of Labor

While our group was originally started with three people, one of our group members has dropped the course, forcing us to change our division of labor. Rachel will generate the data set from the open source repositories and set up the labeling mechanism. Sebastiani will set up an environment in a cloud provider with GPUs and start coding the Network In Network and Inception based models in Keras/Tensorflow and transfer learn layers of a pretrained YOLO onto our problem space.

References

[1] 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.

[2] A. t. Pas, M. Gualtieri, K. Saenko, and R. Platt, "Grasp pose detection in point clouds," *arXiv preprint arXiv:1706.09911*, 2017.