

6.857 Final Project: Milestone 2

Laura Jarin-Lipschitz, Sebastiani Aguirre Navarro and Rachel Holladay

October 12, 2017

1 Algorithms and Evaluation

Our goal is to use machine learning techniques to predict the probability of successfully grasping an unknown object with a robotic arm. Specifically, we will use a convolutional neural network (CNN) to decide whether or not to attempt a simulated robotic grasp, and then quantify the success of our network's decision in simulation. The input to our network will be a point cloud generated from a depth sensor and a candidate grasp, and the output will be a probability of grasp success, which we will threshold to decide whether to make a grasp. The training data will consist of point clouds from the dataset Big Bird [1], and grasp quality labels created by the open source Grasp Pose Generator(GPG) software [2]. The grasping simulation will be performed in robotics simulator Gazebo [3] using the model of the Kinova Jaco Robotic arm and gripper [4].

In recent work using the Big Bird Data set [2], a four-layer CNN with architecture near identical to the LeNet [5] is used. However LeNet was originally developed for handwritten character recognition, and is therefore not necessarily an optimal architecture for this problem. We will explore and compare different CNN architectures to see if we can achieve a higher rate of grasp success as compared to [2]. Given that the dimensions of the images are 1024x1280 in the RGB domain along with 480x640 depth maps, we can test deeper models like Network in Network, Residual Neural Networks or Inception to come up with our own feature extractor or adapt the lower layers of models for object detection that applies similar characteristics as Faster R-CNN, Single Shot Detector or YOLO. The idea is that the lower layers of these object detectors could be adapted to this problem to obtain features related to the object that can be paired with pose features that can be passed to the final classifier layers.

2 Risk Management

One major risk is that our main goal of improving the architecture of the CNN is not achieved. We hope to treat the network in the paper as the baseline and, by exploring other architectures, outperform the network. however there is a

risk that we will not find an architecture with better performance, especially given the difference in our evaluation methods from theirs. However, even if we cannot improve on the percent grasp success in simulation, we hope to explore the effects of different architectures on the types of objects that the network fails to predict correctly.

Another risk we will is our choice to evaluate our CNN's performance in simulation. We made this choice due to time and resource limitations. However, robotics is inherently about a real robot interacting with the physical world. Therefore, there is a part of the project (from the robotics side, not the ML side) that is incomplete, since simulated grasps are highly dependent on the fidelity of the robot model.

3 Division of Labor

Looking forward, we see several immediate tasks. These largely break down into two categories: setting up the infrastructure and baseline and setting up our new architectures. The first, which will be primarily managed by Laura, involves generating the dataset from the open source repositories and setting up the labeling mechanism. In parallel, Sebastiani will lead implementing the pre-processing code to create the dataset representations mentioned in [2]. To jump start progress on the second set of tasks, Sebastiani and Rachel will lead exploring the aforementioned new architectures.

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.
- [3] N. Koenig and A. Howard, "Design and use paradigms for gazebo, an open-source multi-robot simulator," in *IROS*, vol. 3, pp. 2149–2154, IEEE, 2004.
- [4] V. Maheu, P. S. Archambault, J. Frappier, and F. Routhier, "Evaluation of the jaco robotic arm: Clinico-economic study for powered wheelchair users with upper-extremity disabilities," in *ICORR*, pp. 1–5, IEEE, 2011.
- [5] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.