6DOF Grasp Planning by Optimizing a Deep Learning Scoring Function

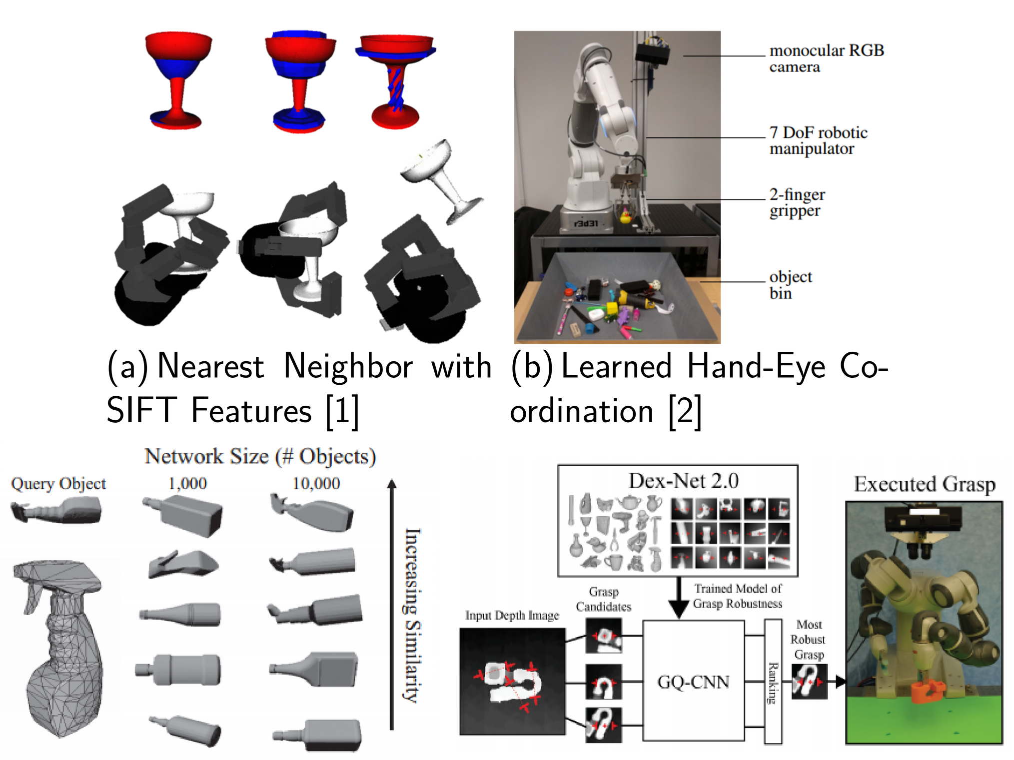
Yilun Zhou & Kris Hauser

Department of Electrical & Computer Engineering
Duke University

Objectives

- Present a method for generating large amount of grasping data using physics simulation;
- Design a deep learning architecture to predict if a hand model can successfully grasp an arbitrary object from an arbitrary pose; and
- Use the learned model to guide grasp planning by optimizing a suboptimal one to an optimal one.

Data-Driven Grasping



(a) Nearest Neighbor with SIFT Features [1] (b) Learned Hand-Eye Co-ordination [2]

(c) Dex-Net 1.0 [3] (d) Dex-Net 2.0 [4]

Figure 1: Previous attempts at data-driven grasping.

Most of the existing works

- Use parallel gripper to approach object from top;
- Generate training data from heuristics (e.g. antipodal points on objects);
- Sample and evaluate random grasps during execution; or
- Do not try to optimize not-so-optimal grasps.

Data Generation

Our data generation method

- Uses full physics simulation to determine feasibility;
- Samples grasp poses uniformly without any priors or heuristics;
- Accommodates any hand model with full 6DOF hand pose; and
- Augments generated data to improve generalization toward novel objects.

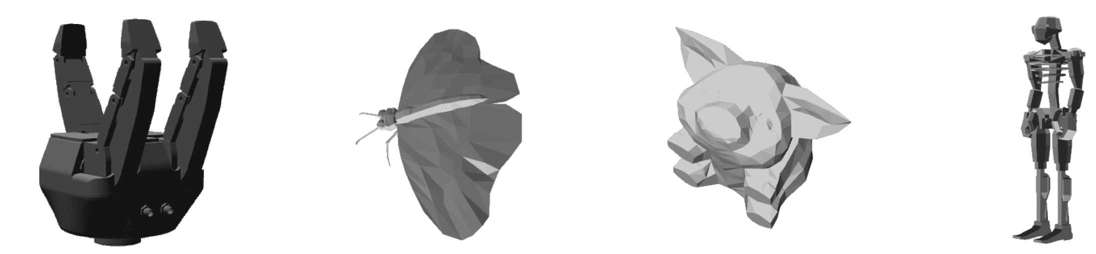


Figure 2: The RobotiQ hand model and objects from Princeton Shape Benchmark [5]

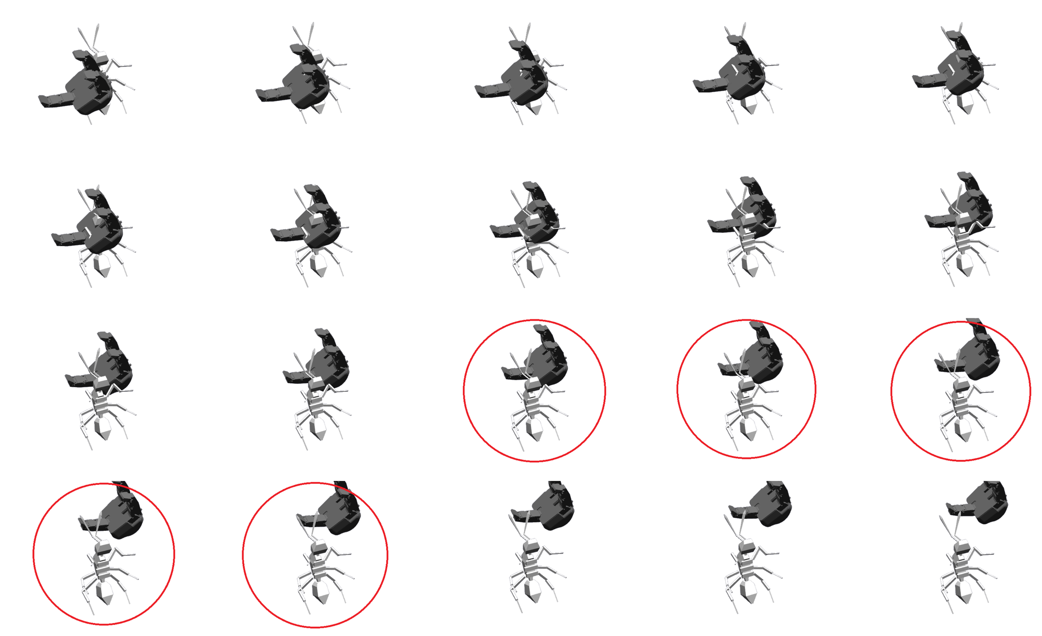


Figure 3: Qualified grasps (circled) are determined geometrically by sliding the hand along a random direction.

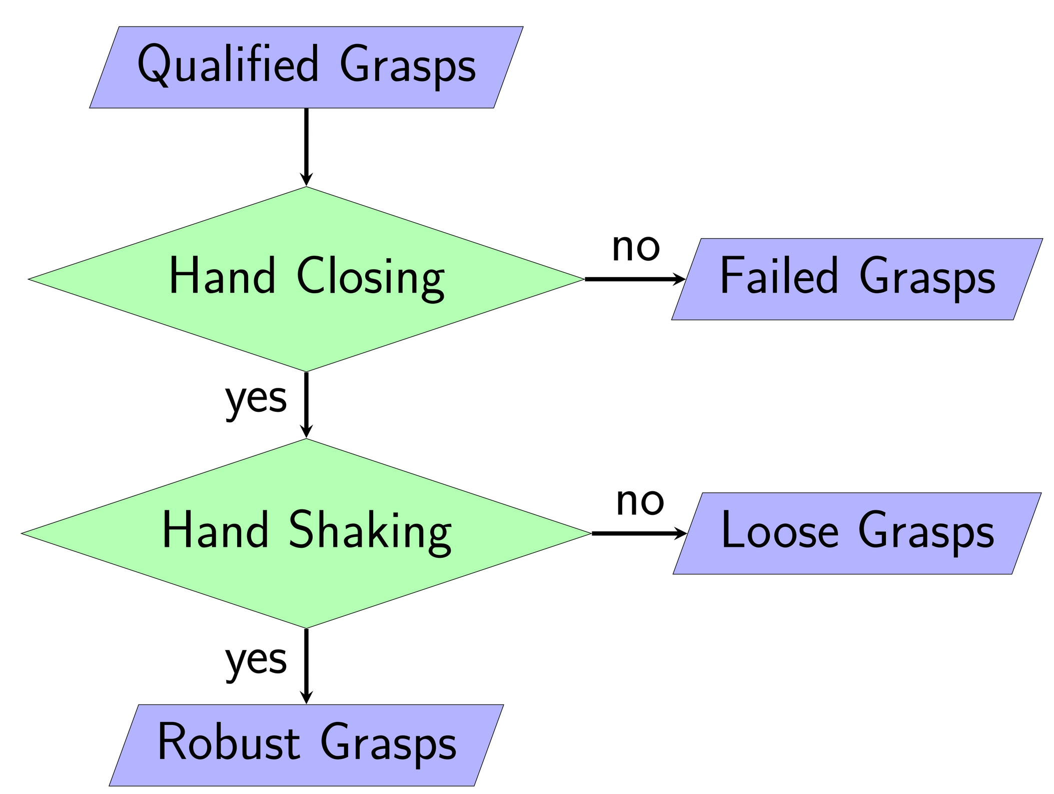


Figure 4: Classification process of qualified grasps

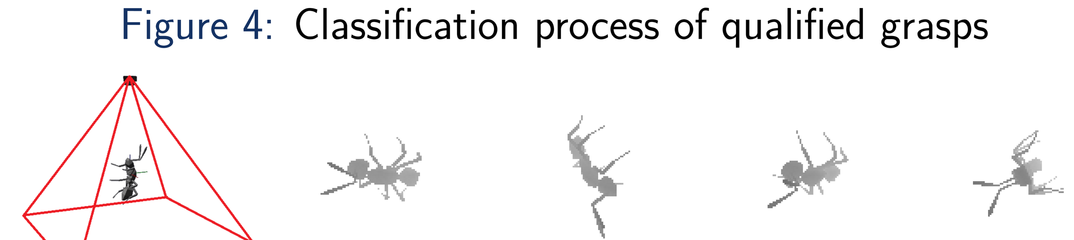


Figure 5: Depth image synthesis and data augmentation.

- 1814 objects
- 442,769 robust grasps
- 1,622,521 loose grasps
- 21,271,561 failed grasps
- 1,000-fold augmentation by synthetic depth images

Grasping is "Multi-Valued" Regression

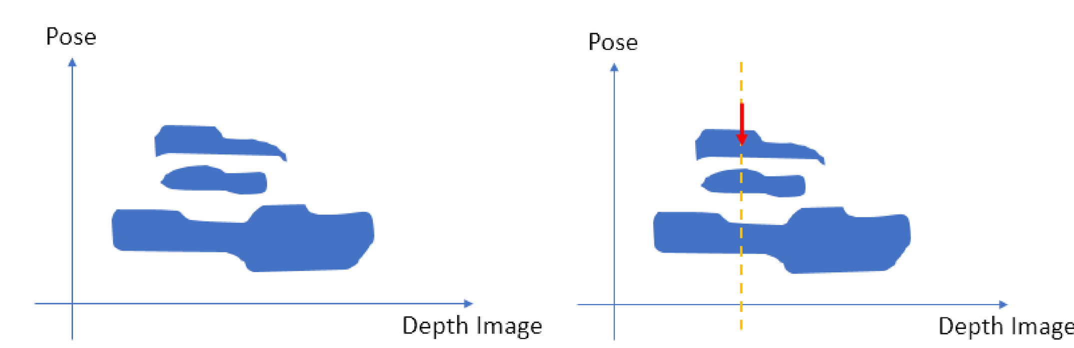


Figure 6: Abstract illustration of the multi-valued problem space (left). We propose learning a classification score and applying a gradient-based control on the pose dimensions (right).

CNN Learning

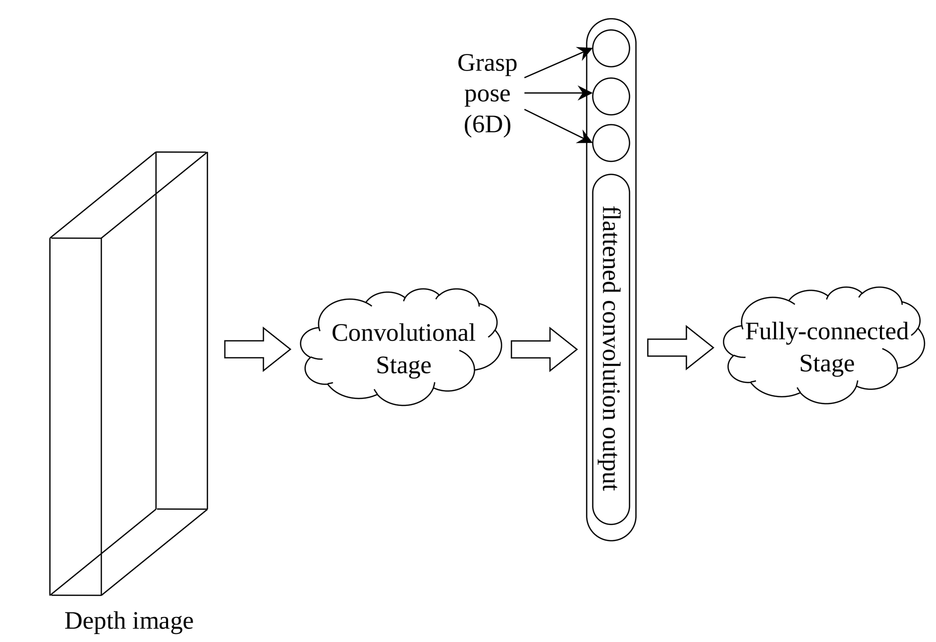


Figure 7: The mix-in CNN architecture used for learning

Object Pose \ Grasp Pose	Train 90%	Test 10%
Train 90%	88.9%	88.7%
Test 10%	88.1%	88.1%

Table 1: Percentage of correct predictions for seen objects. On unseen objects, the percentage of correct predictions is 83.3%.

Gradient-Based Control

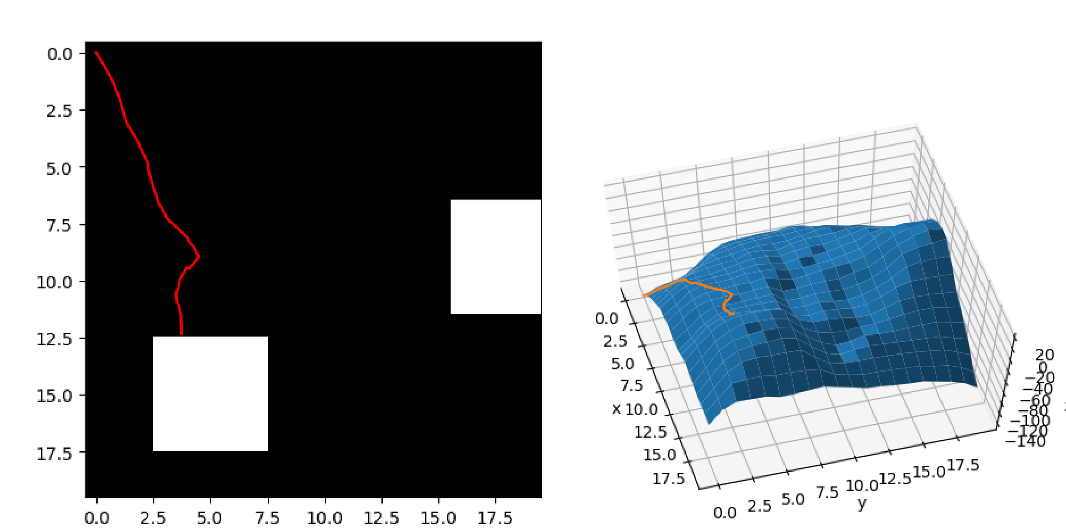


Figure 8: Control on a toy problem, illustrating that CNNs can learn well-behaved scoring functions from classification. Left: the control problem overlayed with optimization trajectory. Right: scoring function is plotted on the z axis.

Objective function used during optimization:

$$f(\mathbf{g}) = x_1(\mathbf{g}, \mathbf{I}) + \frac{\alpha}{\epsilon + distance(hand, object)}.$$

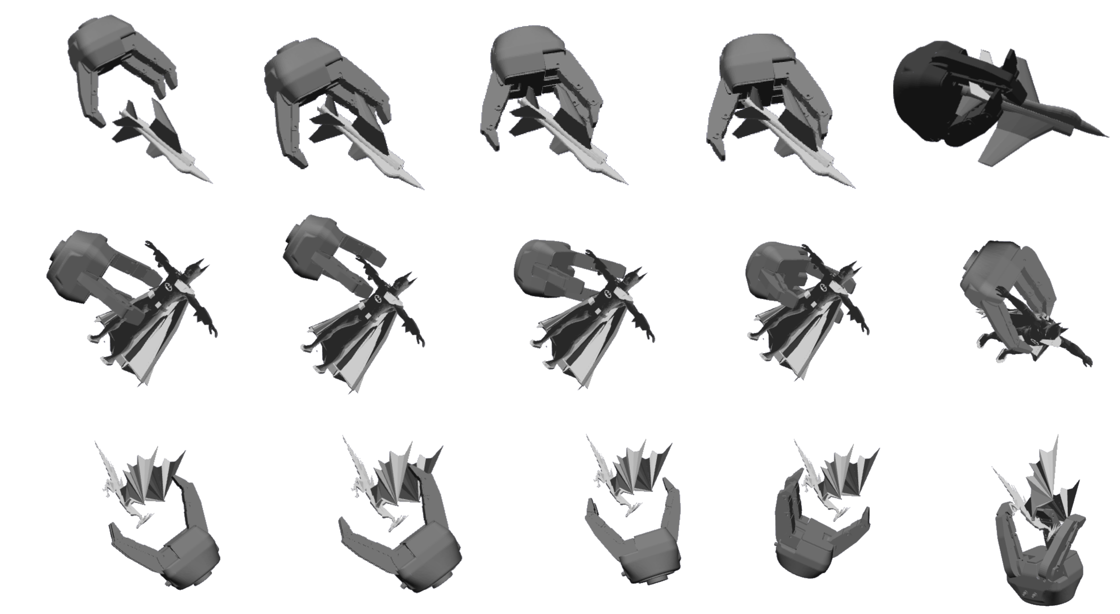


Figure 9: Example sequences for locally optimizing a failed grasp to a robust grasp. After optimization, (top) the hand can successfully grasp the plane by the tail; (middle) the hand moves closer to the Batman and grasps from the side; (bottom) the hand changes to the opposite side of the dragon to better grasp it.

Robust	Loose	Failed
81%	12%	7%

Table 2: Grasp status after optimization. The initial pose is the hand being "pulled away" from a robust grasp pose.

Future Work

- Explore the performance of other CNN architectures for learning [6];
- Explore the performance of other orientation representation (e.g. quaternion); and
- Build the concept of collision avoidance implicitly into the model so that optimization does not require the explicit model.

References

[1] C. Goldfeder, M. Ciocarlie, J. Peretzman, H. Dang, and P. K. Allen, "Data-driven grasping with partial sensor data," in *IROS*, pp. 1278–1283, IEEE, 2009.

[2] S. Levine, P. Pastor, A. Krizhevsky, and D. Quillen, "Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection," *arXiv preprint arXiv:1603.02199*, 2016.

[3] J. Mahler, F. T. Pokorny, B. Hou, M. Roderick, M. Laskey, M. Aubry, K. Kohlhoff, T. Kröger, J. Kuffner, and K. Goldberg, "Dex-Net 1.0: A cloud-based network of 3D objects for robust grasp planning using a multi-armed bandit model with correlated rewards," in *ICRA*, pp. 1957–1964, IEEE, 2016.

[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," in *RSS*, 2017.

[5] P. Shilane, P. Min, M. Kazhdan, and T. Funkhouser, "The Princeton Shape Benchmark," in *Shape Modeling Applications*, pp. 167–178, IEEE, 2004.

[6] Y. Zhou and K. Hauser, "Incorporating side-channel information into convolutional neural networks for robotic tasks," in *ICRA*, IEEE, 2017.