Learning Grasp Evaluation Models Using Synthetic 3D Object-Grasp Representations

Minh Nguyen[1], Paul G. Plöger, Alex Mitrevski[1], and Maximilian Schöbel[1]

Hochschule Bonn-Rhein-Sieg, Grantham-Allee 20, 53757 Sank Augustin, Germany
minh.nguyen@smail.h-brs.de,
{paul.ploeger,aleksandar.mitrevski,maximilian.schoebel}@h-brs.de

Abstract. This project considers the problem of generating data for training grasp evaluation models. Recent advances are reviewed for four main aspects most relevant to labeled grasp data synthesis, namely feature extraction from perceptual data, object-grasp representation, grasp evaluation techniques, and data generation techniques. From this review, one may conclude that while data synthesis for learning a grasp evaluation model is promising, recent approaches are either limited by difficulties in collecting large-scale human grasp experience, or by the shortcomings of using analytical metrics to label generated data. Additionally, a completed object grasping pipeline is integrated, from object detection to grasp pose detection and grasp execution. Two set of experiments are performed on the Toyota Human Support Robot for two pose estimation methods using this grasping pipeline. The pipeline proves reliable and fast enough for performing the experiments, being able to execute 20 grasps per object without interruption. While further extension and optimization are needed, the pipeline enables directly examining and comparing more advanced grasp planning methods in the future.

Keywords: Grasp learning · Data synthesis.

1 Introduction

1.1 Motivation

Robot grasping with multi-fingered robotic hands is a challenging problem, and finding a grasp planning solution which resembles humans' grasps in dexterity and robustness is still an area of active research.

1.2 Use Case

- This project focuses on grasping tasks relevant to the Robocup@Home [1] competition. These tasks involve objects which can be commonly found in a domestic environment, some of which can be seen in figure 1.
- The experiments will be conducted on the Human Support Robot (HSR) from Toyota [2].

[1] http://www.robocupathome.org
[2] https://www.toyota-global.com/innovation/partner_robot/robot/

Fig. 1: Typical objects in the Robocup@Home competition [14].

1.3 Overview of robotic grasping research

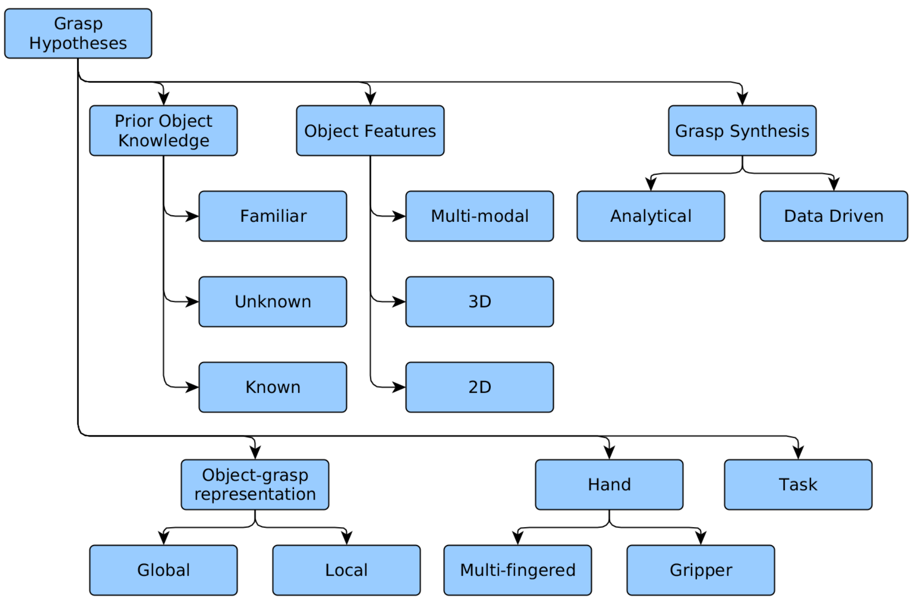


Fig. 2: Aspects which may influence generation of grasp hypotheses [2].

Sahbani et al. [19] classify grasp synthesis into analytical and empirical approaches:

– Analytical methods consider mechanical properties of the contact points between the gripper's fingers and an object [18,19,21]: disturbance resistance, dexterity, equilibrium and stability.

– Empirical methods rely on some form of grasp experience to synthesize candidates. Bohg et al. [2] organize them based on how much information is assumed about the object: whether they are known, familiar or completely unknown to the robot.

2 State of the Art

2.1 Feature extraction from perceptual data for grasping

– Several approaches Convolutional Neural Networks (CNN) to combine depth and RGB features [5,8,16]
– Some approaches predict regions occluded in the RGB-D point clouds [17,22] or reconstruct objects from the incomplete view of a single RGB-D camera [1,23].

2.2 Object-grasp representation

– Based on global features: can use CNN on RGB-D image of object [13] or eigengrasps [3,6].
– Based on local features: extract features from the rectangle/cuboid regions corresponding to a two-fingered gripper [4,11,9], or local template grids at points of contact [10].

2.3 Grasp evaluation

– Analytical grasp metrics: focus on form/force closure properties, generally via analyzing the grasp matrix G, the force wrench space (FCS), the grasp wrench space (GSP), or the hand-object Jacobian J.
– Analytical metrics can be extended to consider task affordance via limiting the analysis only to movements relevant to performing a specific task.
– Learning to predict grasp quality: table 1 summarize the most prominent approaches.

2.4 Generating data for grasp success prediction

– Data synthesis for robot grasping
– Data augmentation

3 Methodology

3.1 Object detection

Integrate Single-Shot Multibox Detector (SSD) for object detection (figure 3)

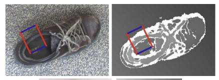
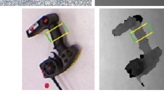
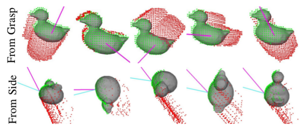
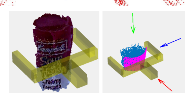
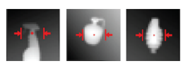
Method	Object-grasp representation	Feature extraction & learning model	Data generation
[9]		Histogram of hand-crafted filters; Model: SVM.	Rectangles manually annotated.
[11]		Auto-encoders to initialize weights, structured regularization to combine depth and RGB data; Model: MLP.	Extension of the dataset from [9] (above).
[10]		RGB rendering of "template grids"; Model: LeNet CNN	Quality of grasps are calculated in simulation for object meshes, verified via crow-sourcing.
[7]		Filters of cuboid regions projected onto 3 orthogonal planes, creating 15 channels; Model: LeNet CNN.	Quality of grasps are calculated for object meshes using force-closure
[13]		Depth images cropped and aligned to gripper; Model: CNN combined with single-layer NN.	Quality of grasps are calculated for object meshes using a variant of ϵ-metric from [24]

Table 1: Five recent empirical approaches to grasp quality prediction

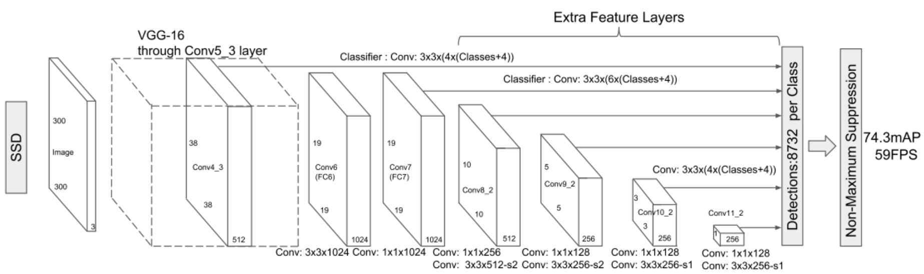


Fig. 3: SSD architecture [12].

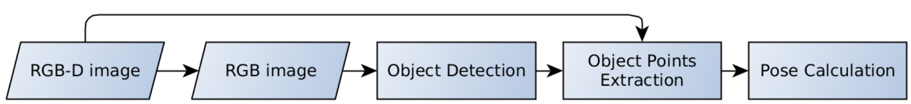


Fig. 4: Flowchart of the baseline method for grasp experiments.

3.2 Pose estimation and grasping

Baseline method

- assume the grasp approach to be along the x-axis of the robot
- extract object points using image detection result
- estimate object positions in the `base_link` coordinate frame
- estimated y- and z-coordinates are mean of the object points

– estimated x-coordinate is either the minimum or mean of the object points

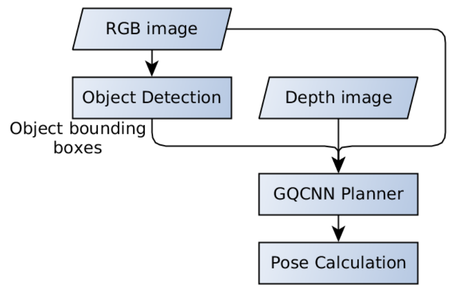


Fig. 5: Flowchart of the integrated GQCNN grasp planner [13].

Grasp Quality CNN (GQCNN)

– GQCNN model [13] is trained on Dex-Net 2.0 dataset to predict grasp robustness,
– In the original setup, the grasp is assumed to be from above and perpendicular with the table,
– Grasp approach vector is assumed to align with camera axis,
– Not reliable enough for performing grasp experiments.

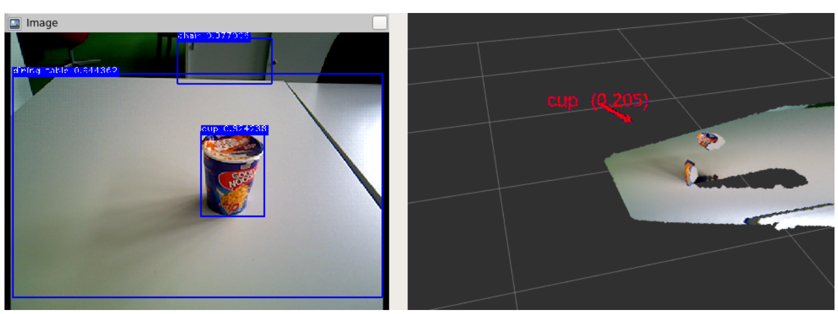


Fig. 6: A successful GQCNN grasp plan. Arrow and number on the right indicate grasp pose and quality returned from the GQCNN planner.

4 Experiments

4.1 Experimental Setup

– Two sets of experiments are performed for the variances of the baseline method described in slide
– Before each grasp, the robot is moved to a marked position facing the dining table.
– Objects further than 90cm in x-axis and lower than 75cm in z-axis are ignored.
– We use the SSD model pre-trained on the COCO dataset.
– Arm collisions and object slips are counted as failures.
– Only one of the objects in figure 7 is grasped at a time.

(a) Styrofoam ball (b) Duct tape (c) Noodles box (d) Salt container

Fig. 7: Objects selected for the experiments.

4.2 Results

Object	Mean x		Minimum x	
	Success	Failure	Success	Failure
Salt	17	3	16	4
Ball	8	12	15	5
Noodle box	16	4	12	8
Duct tape	7	13	13	7

Table 2: Results of the grasp experiments. On the left are results from using the mean x coordinates for estimating the grasp pose, and on the right are results from using the min coordinates along the x-axis.

5 Conclusion

5.1 Contributions

The first contribution of this work is a detailed review of recent advances in aspects most relevant to generating data for training a grasp evaluation models, namely feature extraction from perceptual data, object-grasp representation, grasp evaluation metrics, and data generation techniques. Additionally, five recent, prominent approaches to data synthesis for grasp evaluation are examined, and their solutions for each of the four aspects mentioned above are summarized in table 1.

The second contribution of this project is the implementation of a full grasping pipeline, from perceiving objects to grasp execution, in collaboration with another Research and Development project by Padalkar [15]. Two pose estimation methods are implemented, serving as baselines for experimenting and comparing with more advanced grasp planning techniques.

The review of recent approaches to grasp data synthesis demonstrates their limitations either in dataset size or by using theoretical approaches to generate data labels, suggesting possible extensions and improvements with larger human grasp experience database [20] or more advanced feature extraction methods [23].

5.2 Future work

- Grasp execution can be extended for grasping from multiple directions.
- SSD model for detection can be fine-tuned for RoboCup objects.
- Grasp pipeline should be tested on the Care-o-bot.
- More advanced method can be implemented to find better pose estimates using the object points.
- Surface normal can be calculated for a better approach vector estimate.
- Approaches in table 1 can be integrated and extended with new human grasp database [20].
- Techniques introducing task awareness into data generation and labeling can also be examined.

References

1. Bohg, J., Johnson-Roberson, M., Len, B., Felip, J., Gratal, X., Bergstrm, N., Kragic, D., Morales, A.: Mind the gap - robotic grasping under incomplete observation. In: Zheng, Y.F. (ed.) 2011 IEEE International Conference on Robotics and Automation. pp. 686–693 (May 2011). https://doi.org/10.1109/ICRA.2011.5980354
2. Bohg, J., Morales, A., Asfour, T., Kragic, D.: Data-driven grasp synthesis: A survey. IEEE Transactions on Robotics **30**(2), 289–309 (April 2014). https://doi.org/10.1109/TRO.2013.2289018
3. Ciocarlie, M.T., Allen, P.K.: Hand posture subspaces for dexterous robotic grasping. The International Journal of Robotics Research **28**(7), 851–867 (2009). https://doi.org/10.1177/0278364909105606, https://doi.org/10.1177/0278364909105606

4. Detry, R., Ek, C.H., Madry, M., Piater, J., Kragic, D.: Generalizing grasps across partly similar objects. In: Bicchi, A. (ed.) 2012 IEEE International Conference on Robotics and Automation. pp. 3791–3797 (May 2012). https://doi.org/10.1109/ICRA.2012.6224992
5. Eitel, A., Springenberg, J.T., Spinello, L., Riedmiller, M., Burgard, W.: Multimodal deep learning for robust rgb-d object recognition. In: Burgard, W. (ed.) 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 681–687 (Sept 2015). https://doi.org/10.1109/IROS.2015.7353446
6. Goldfeder, C., Allen, P.K.: Data-driven grasping. Autonomous Robots **31**(1), 1–20 (Jul 2011). https://doi.org/10.1007/s10514-011-9228-1, https://doi.org/10.1007/s10514-011-9228-1
7. Gualtieri, M., ten Pas, A., Saenko, K., Platt, R.: High precision grasp pose detection in dense clutter. In: et al., W.B. (ed.) 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 598–605 (Oct 2016). https://doi.org/10.1109/IROS.2016.7759114
8. Gupta, S., Girshick, R., Arbeláez, P., Malik, J.: Learning rich features from rgb-d images for object detection and segmentation. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) Computer Vision – ECCV 2014. pp. 345–360. Springer International Publishing, Cham (2014)
9. Jiang, Y., Moseson, S., Saxena, A.: Efficient grasping from rgbd images: Learning using a new rectangle representation. In: Bicchi, A. (ed.) IEEE International Conference on Robotics and Automation (ICRA). pp. 3304–3311 (May 2011). https://doi.org/10.1109/ICRA.2011.5980145
10. Kappler, D., Bohg, J., Schaal, S.: Leveraging big data for grasp planning. In: Okamura, A. (ed.) IEEE International Conference on Robotics and Automation (ICRA). pp. 4304–4311 (May 2015). https://doi.org/10.1109/ICRA.2015.7139793
11. Lenz, I., Lee, H., Saxena, A.: Deep learning for detecting robotic grasps. The International Journal of Robotics Research **34**(4-5), 705–724 (2015). https://doi.org/10.1177/0278364914549607, https://doi.org/10.1177/0278364914549607
12. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C.: Ssd: Single shot multibox detector. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) Computer Vision – ECCV 2016. pp. 21–37. Springer International Publishing, Cham (2016)
13. Mahler, J., Liang, J., Niyaz, S., Laskey, M., Doan, R., Liu, X., Ojea, J.A., Goldberg, K.: Dex-net 2.0: Deep learning to plan robust grasps with synthetic point clouds and analytic grasp metrics. CoRR **abs/1703.09312** (2017), http://arxiv.org/abs/1703.09312
14. Matamoros, M., Rascon, C., Hart, J., Holz, D., van Beek, L.: Robocup@home 2018: Rules and regulations. http://www.robocupathome.org/rules/2018_rulebook.pdf (2018)
15. Padalkar, A.: Dynamic motion primitives. Tech. rep., Hochschule Bonn-Rhein-Sieg (2018)
16. Porzi, L., Bul, S.R., Penate-Sanchez, A., Ricci, E., Moreno-Noguer, F.: Learning depth-aware deep representations for robotic perception. IEEE Robotics and Automation Letters **2**(2), 468–475 (April 2017). https://doi.org/10.1109/LRA.2016.2637444
17. Qi, C.R., Su, H., Niener, M., Dai, A., Yan, M., Guibas, L.J.: Volumetric and multi-view cnns for object classification on 3d data. In: Agapito, L.,

Berg, T., Kosecka, J., Zelnik-Manor, L. (eds.) 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5648–5656 (June 2016). https://doi.org/10.1109/CVPR.2016.609

18. Roa, M.A., Suárez, R.: Grasp quality measures: Review and performance. Autonomous Robots **38**(1), 65–88 (Jan 2015). https://doi.org/10.1007/s10514-014-9402-3, http://dx.doi.org/10.1007/s10514-014-9402-3
19. Sahbani, A., El-Khoury, S., Bidaud, P.: An overview of 3d object grasp synthesis algorithms. Robotics and Autonomous Systems **60**(3), 326–336 (Mar 2012). https://doi.org/10.1016/j.robot.2011.07.016, http://dx.doi.org/10.1016/j.robot.2011.07.016
20. Saudabayev, A., Rysbek, Z., Khassenova, R., Varol, H.A.: Human grasping database for activities of daily living with depth, color and kinematic data streams. Scientific Data **5**, 180101 (May 2018), http://dx.doi.org/10.1038/sdata.2018.101, data Descriptor
21. Shimoga, K.: Robot grasp synthesis algorithms: A survey. The International Journal of Robotics Research **15**(3), 230–266 (1996). https://doi.org/10.1177/027836499601500302, https://doi.org/10.1177/027836499601500302
22. Su, H., Maji, S., Kalogerakis, E., Learned-Miller, E.: Multi-view convolutional neural networks for 3d shape recognition. In: Ikeuchi, K., Schnrr, C., Sivic, J., Vidal, R. (eds.) 2015 IEEE International Conference on Computer Vision (ICCV). pp. 945–953 (Dec 2015). https://doi.org/10.1109/ICCV.2015.114
23. Varley, J., DeChant, C., Richardson, A., Ruales, J., Allen, P.: Shape completion enabled robotic grasping. In: Maciejewski, T. (ed.) 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 2442–2447 (Sept 2017). https://doi.org/10.1109/IROS.2017.8206060
24. Weisz, J., Allen, P.K.: Pose error robust grasping from contact wrench space metrics. In: Bicchi, A. (ed.) 2012 IEEE International Conference on Robotics and Automation. pp. 557–562 (May 2012). https://doi.org/10.1109/ICRA.2012.6224697