2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
Madrid, Spain, October 1-5, 2018

Learning Robotic Grasping Strategy Based on Natural-Language Object Descriptions

Achyutha Bharath Rao, Krishna Krishnan, and Hongsheng He*

Abstract— **Given the description of an object's physical attributes, humans can determine a proper strategy and grasp an object. This paper proposes an approach to determine grasping strategy for an anthropomorphic robotic hand simply based on natural-language descriptions of an object. A learning-based approach is proposed to help a robotic hand learn suitable grasp poses starting from the natural language description of the object. Object features are parsed from natural-language descriptions by using a customized natural-language processing technique. The most likely grasp type for the given object is learned from the human grasping taxonomy based on the parsed features. The grasping strategy generated by the proposed approach is evaluated both by simulation study and execution of the grasps on an AR10 robotic hand.**

I. Introduction

A five-digit hand configuration with an opposable thumb in hominids is generally considered to be one of the most important evolutionary developments that led to human advancement. We humans, in turn, have designed the world around us for convenient grasping and manipulation by our hands. So it is not surprising that, when you ask a blindfolded person to pick up an object by describing its features such as shape, size and mass; she can choose a feasible grasp in an instant.

However, it has been very challenging to codify this human grasping skill and transfer the ability to robots. Recent research trends are focused on learning-based approaches - but most of the work is focused on two or three-fingered robotic clamps [23], [28]. The research of grasp planning for an anthropomorphic robotic hand is challenging and deserves more effort [3], [4]. In this study, we propose to learn grasping by a humanoid robotic hand using only the natural-language descriptions of an object - something akin to asking a blindfolded person to grasp an object.

People internalize dexterity by repeated interaction with objects. Our approach is to learn robotic grasps by emulating this human learning process. The framework of the proposed approach is shown in Fig. 1. A key idea of this study is to discritize the 10 degrees-of-freedom (DOF) joint configuration space into 6 distinct human grasp type subspaces, whereby only the grasp type and scale need to be learned. This approach reduces the problem dimensionality and renders it into a multi-class classification problem of selecting one feasible human like grasp. The success of the approach is evaluated by scoring the predicted grasps against the human labeled grasps and also by executing the grasps on an AR10 robotic hand with new unfamiliar set of objects.

Achyutha Bharath Rao and Krishna Krishnan are with the Department of Industrial Systems and Manufacturing Engineering, Wichita State University, Wichita, KS, 67260, USA axbharathrao@shockers.wichita.edu, krishna.krishnan@wichita.edu.

Hongsheng He is with the Department of Electrical Engineering and Computer Science, Wichita State University, Wichita, KS, 67260, USA hongsheng.he@wichita.edu.

*Correspondence should be addressed to Hongsheng He, hongsheng.he@wichita.edu.

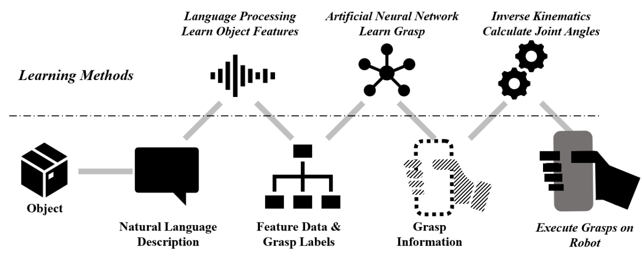


Fig. 1. Framework of the proposed approach.

II. Related Work

A. Understanding Human Grasps

Most of the efforts in understanding grasps [6], [8] have been to breakdown the human grasping actions into discrete classes. A structured classification of grasps is discussed in [6] based on object shapes and task requirements. Recently, a more comprehensive version of the grasp taxonomy has been developed [8] by de-coupling them from the object shapes and the tasks being performed.

A neuroscience based approach to simplify and understand human grasps is proposed in [25]. The study reports that hand posture can be decomposed into very few general configurations and that the finer adjustments can be achieved by superposition of such grasp poses. The research in [5] is built on this concept and has proposed a method of using 'eigengrasps' to reduce dimensionality of grasps. Reducing dimensionality is a necessary step to make the problem of learning grasps tractable. In our study we have used a different approach to achieve the same goal.

B. Learning Robotic Grasps

Robotic dexterity has been a difficult goal for a while and multiple approaches have been proposed to help robots master the grasping skill. Earlier methods involved analytical approaches to calculate object affordances and contact forces to determine grasp successes [19]. Expert systems related approaches have attempted to logically codify the grasp choices for a set of object features [6], [29]. But the sheer number of variations of human grasps limit such approaches to few narrow applications. Recent proposals have therefore, focused on learning methods [16], [10], [12], [1], [27], [17], especially application of deep learning methods to learn grasps [14], [13], [24], [26]. Learning techniques have been extensively used to solve object recognition, pose estimation, grasp planning and execution [15], [24]. Most such studies [18] were focused on object recognition using images or 3D point clouds.

This paper subscribes to the idea that robotic dexterity can be improved by emulating human grasps and the means to achieve this is by using learning methods. There have, of course, been many studies which are pursuing such hypothesis using probabilistic reasoning [27], [17], machine learning methods. Heinemann *et al*. [9] used human grasping as the basis to train a three-fingered clamp to use the contact between the surfaces and the hand to achieve grasps.

C. Natural Language Processing

Specific to robotics, natural language descriptions to understand object affordances, have been studied but mostly in the context of complementing machine vision [2] and to recognize objects [21]. Wang *et al.* [30] discussed the use of fixed patterns for natural-language parsing and extracting attributes but the focus was on object recognition. Farhadi *et al.* [7] used object descriptions for the purpose of object identification. Our goal here is a novel one, in that, we are trying to parse specific attributes from natural-language descriptions for the purpose of learning to grasp the object. To the best of our survey, this paper is the first work that uses natural-language descriptions to aid robotic grasping.

III. Learning Grasping Strategy

The goal of our study is to learn a mapping that takes in a natural-language description of a set of physical features F of an object and outputs a grasp type G.

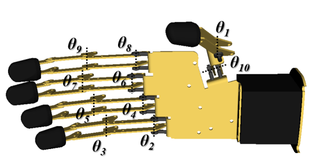


Fig. 2. AR10 Robotic Hand with 10 Degrees of Freedom $[\theta_1, \theta_2, ...\theta_{10}]$.

A grasp posture is defined by the joint angles of the fingers. In this paper, we will use the AR10 robotic hand that has 10 DOF, as shown in Fig. 2. Therefore, each grasp posture G is a 10 dimensional vector of joint angles, defined as

$$G = [\theta_1, \theta_2, ...\theta_{10}] \in \mathbb{R}^{10}$$

Learning ten labels for each grasp is nontrivial. In this paper, we map this 10-dimensional space to a 2-dimensional subspace given by

$$g = [h_i, \alpha]$$

where $h_i \in h_1, h_2, .., h_k$ represents one of the human grasp types and $\alpha \in \mathbb{R}$ is a scalar which determines the size of the grasp. Each $h_i = [\theta_{i1}, \theta_{i2}, ...\theta_{i10}]$ is a unique combination of joint angles representing one of the human grasps with θ_{ij} chosen such that h_i mimics a particular human grasp type from the grasp taxonomy.

The g then maps to G as

$$G_i = \alpha h_i$$

Therefore, we can define a range of grasps using just two parameters, namely the human grasp type h_i and the scalar α. So the problem now can be re-stated as learning a mapping between the set of object features F to $g_i = [h_i, \alpha]$.

A. Grasp Taxonomy

Most studies [6], [20], [11] attempting to understand and codify human grasps have come to conclusion that human grasp choice is a function of object affordances and the task requirements. For one specific object/task combination, there could be multiple grasp choices possible. However, human grasp choices do tend to cluster [25], [8] when studied over a large set of objects.

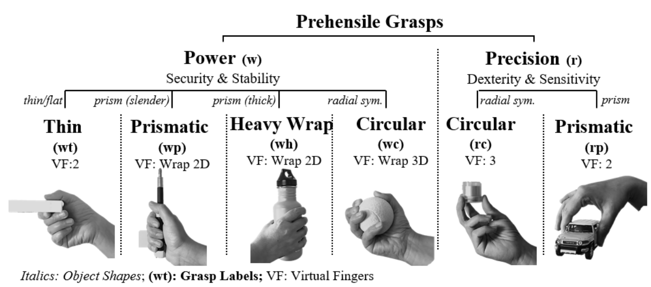


Fig. 3. Human Grasp Taxonomy derived from [6].

Given this understanding, our goal is to select one of the feasible grasp types for a given object. Due to the limitations posed by a 10 DOF robotic hand it was decided to use the simpler grasp taxonomy presented by [6] for the purpose of our study. We use the six higher level classifications from [6] and derive the finer adjustments by combining these six grasp types with the scalar α. The chosen human grasp classification and the nomenclature for each grasp is shown in Fig. 3, where the prefixes $'w'$ and $'r'$ stand for Power and Precision grasps [20], [6].

B. Object Attributes of Interest

Human grasp strategies depend on various factors including object shape, size, weight, texture, relative position, orientation, function and so on [20]. In human grasping, adapting for variations in position and orientation of an object will require the use of the whole arm and not just the hand. Since the focus of this study is limited to the robotic hand, it was decided to keep the position and orientation constant. Based on the research we define the feature set F of an object as shown on Table I.

So our goal is to map this feature set F to the grasp g,

$$F = [a, b, c, m, s, r, mt] \rightarrow g = [h_i, \alpha]$$

TABLE I

DESCRIPTION OF OBJECT FEATURE SET F.

Feature	Description	Value Range
a,b,c	Dimensions (cm) along orthogonal directions [8]	$a,b,c \in \mathbb{R}$ and $a \geq b \geq c$
m	Mass (grams)	mass $\in \mathbb{R}$
s	Shape classification per [6]	{thin, compact, prism, long, radial}
r	Rigidity of the object	{rigid, squeezable, floppy}
mt	Simplified material type description of the object.	{fabric, glass, metal, paper, plastic, rubber, wood, other}

C. Translating Object Descriptions to Features

Humans tend to describe objects by stating approximate dimensions and salient features such as "it is about ten centimeters long and two centimeters in diameter, weighs about hundred grams, is made of plastic.". Our attempt is to use customized Natural Language Processing (NLP) techniques to estimate the object features F that are needed to learn the grasp types.

$$l_1 \rightarrow F_1 = [a_1, b_1, c_1, m_1, s_1, r_1, mt_1]$$

where l_1 is a training datum containing description of a sample object.

Using a combination of lemmatizing, parts-of-speech tagging and chunking, we extract any available quantitative and qualitative descriptors of the object(s). We look for expressions chunks such as "two centimeters long","made of plastic" or "very heavy", and then we create a chunk-tree using regular expressions as shown in Fig. 4. In case of missing dimensions, we perform data imputation using a rule based approach of estimating the missing dimension based on the other available dimensions. The rule itself was derived from the priors in the data. The success of this NLP model is evaluated by scoring the parsed values with the measured/labeled values and the scores are used improve the NLP extractor.

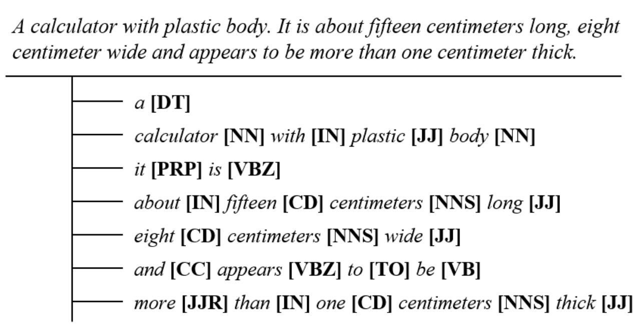


Fig. 4. Example of the chunk tree. The input statement at the top is parsed into chunks as shown in the tree using regular expressions.

D. Grasp Definition

The output labels obtained from the NLP need to be used for estimating a suitable grasp. To accomplish this, we need to learn to select suitable grasp type using the experimentally compiled grasp dataset. The grasp dataset consists of two labels namely the Grasp Type and Grasp Dimension against each object in the dataset compiled experimentally as described in section IV.

Grasp type labels are based on the grasp classification shown in Fig. 3. The grasp type h_i is drawn from the set of human grasp primitives as shown

$$h_i \in wt, wp, wh, wc, rp, rc$$

Grasp Dimension corresponds to a,b,c or their combinations, around which the grasp closure occurs. Labels similar to this are used in [8] and is useful for deciding the extent of hand closure while modeling robot forward/inverse kinematics.

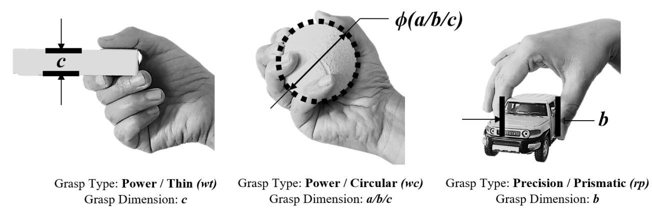


Fig. 5. Sample illustrations of grasp type and dimension(size).

Grasp dimension d_i is given by

$$d_i \in [a, b, c, ab, bc, ac, abc]$$

Evidently, the grasp choice and grasp dimension are interdependent i.e. certain grasps prefer certain grasp dimensions. But for certain object-grasp combinations, there could be multiple grasp dimensions associated with same grasp. Such variations were often seen with objects whose a/b or b/c ratios were close to1. Instead of forcibly pairing grasps and dimensions, we decided to treat the two labels h_i and d_i as independent. The rationale was that any latent pairing tendencies between grasp-types and grasp-dimensions will manifest within the learned model. This assumption can be revisited in future studies.

E. Learning Grasps

As outlined in Fig. 6, the model consists of two separate neural networks – a grasp type selection $\tau(F)$ and a grasp dimension selection model $\delta(F)$. $\tau(F)$ models a probability distribution $p(h_i|F)$ over the six possible grasp types for the given object affordances that corresponds to the normalized probability of a human selecting such a grasp. Similarly, $\delta(F)$ models a probability distribution $p(d_i|F)$ of dimensions(s) along which the object is held. The output of $\tau(F)$ and $\delta(F)$ are softmax over the possible discrete values of h_i and d_i.

The neural network model consisted a hidden layer with 75 units activated by ReLU. In order to tune L_2, grid search was employed and $alpha = 3.6$ was chosen. The model was optimized using standard back-propagation and scored with a 4-fold cross-validation to learn the grasp type and grasp-dimension.

At the end of this step we have successful mapping from $F=[a,b,c,m,s,r,mt]$ to $g'=[h_i,d_i]$. Next step is to convert g' to g given by $[h_i,\alpha]$.

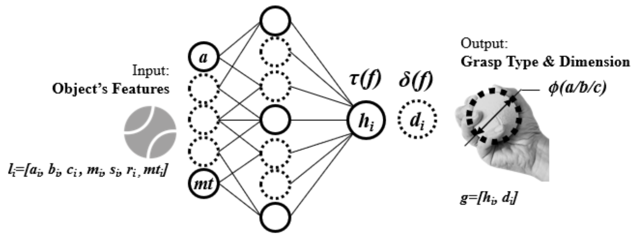


Fig. 6. Grasp Selection Neural Network Classification Model

F. Inverse Kinematics

The grasp dimension d_i, thus obtained, needs to be converted to the grasp size p which is essentially, the distance between the virtual fingers [6] of a particular grasp. For most grasps, the grasp size p corresponds to the dimension(s) a, b, or c along which the object is held. For grasps which are held along two or more dimensions (e.g. heavy wrap or circular), the grasp size is derived from grasp dimension based on the shape of the object such as the diagonal distance of a prismatic object or diameter of a cylindrical or spherical object. Distance between virtual fingers, p, is a function of the grasp type h_i and scalar α.

$$p = f(\alpha.h_i) \text{ where } h_i = [\theta_1, \ldots \theta_{10}] \text{ and } \alpha \in \mathbb{R}$$

For a given grasp, h_i is fixed. Within the limited range of motion of each grasp type, p is assumed to be linearly related to α, and therefore

$$p = w_{1i}\alpha + w_{0i}$$

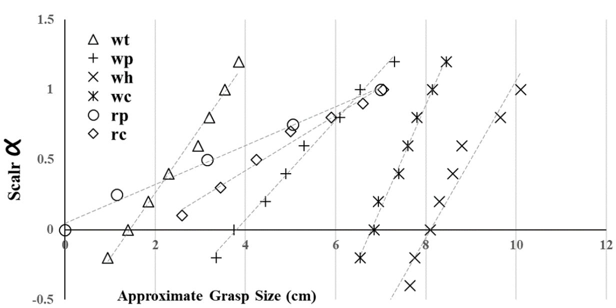


Fig. 7. Inverse Kinematics Charts. Data generated by varying α and measuring Grasp Size for each human grasp type. Using this data, we derive a linear Inverse Kinematics model for each grasp.

To generate the grasp curves to fit the linear model, we first manually tune the joint angles to match a grasp type, say wp (Power-Prismatic). We then scale the all the joint angles by a scalar α_x and measure the distance p_x between the virtual fingers of the grasp. We capture data for various values of α_x. Using the data, we learn the parameters of a linear model mapping α to p for each grasp type. We learn the parameters w_1 and w_0 by fitting a linear model. Now, we can use w_0 and w_1 in forward kinematics (p given α) and inverse kinematics (α given p) calculations.

For a new object instance, once we have the learned grasp types and sizes from earlier models, we use the inverse kinematics model to configure the robotic hand to the desired grasp and perform the grasp action.

IV. Experiments

A. Data Collection

Objects Dataset: A library of 100 objects of everyday use were compiled to create the objects data-set. The objects were chosen such that they span a wide range of values of the features e.g. objects with dimensions spanning from very small to very large etc. The data-set consisted of dimension measurements, mass and manually labeled rigidity, material & shape classifications. For generating the natural language descriptions, we arbitrarily split the data in half and engaged two individuals and asked them to write descriptions for each of the objects. The subjects were instructed as to purpose of the description and to include descriptions of as many physical features as possible. A ruler and kitchen weighing scale were provided to assist the individuals to better estimate dimensions and mass.

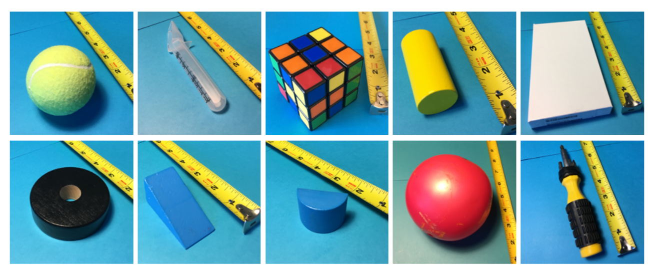
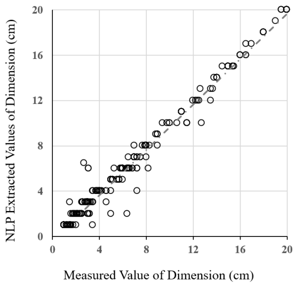


Fig. 8. A sample of 10 objects that constituted the object dataset.

Grasp Dataset: The goal here was to map each object to a grasp type label. There can be more than one grasp type for 'holding' a given object. We, therefore, replicated the experiment with two different subjects. We instructed the individuals to try diverse ways of grasping the object and settle on one which they believe to be secure enough to pick and translate the object. Variations in grasp choices between individuals were resolved by discussing with the subjects and by repeating the grasps. The final data set was chosen to have one set of labels for grasp type and grasp dimension / size.

B. Learning Phase

Natural Language Processing: Language processing was performed as described in section III-C. To score the numerical values we used Least Squares Regression model. The final model was able to fit with an R^2 of 0.98 overall for dimensions and 0.87 for mass estimations. The regression fit for dimension estimations is shown in Fig. 9. Categorical labels for $material, shape, rigidity$ were scored as well. Fig. 9 shows the scoring matrix for $'material'$. The primary source of errors in this step are approximations of quantities such as mass and dimensions. The second source of errors are introduced by the inaccuracies in parsing natural language. There is scope for improving the parsing algorithm in future.



Labeled ↑ \ Predicted →	*fabric*	*glass*	*metal*	*paper*	*plastic*	*rubber*	*wood*	*other*	
fabric	2				1			1	4
glass		4						2	6
metal			18		2			2	22
paper				13					13
plastic					40			1	41
rubber						3			3
wood							7	1	8
other								3	3
	2	4	18	13	43	3	7	10	

Fig. 9. Left: Regression curve showing NLP parsed and measured values of dimensions. Right: A scoring map of predicted vs. labeled values of *material* type.

Learning Grasps: The parsed object features were fed into the neural network classifier [22] to learn grasp selection strategy. The cross-validation scores for both the grasp type and grasp dimension converged independently to an accuracy of 0.79 ± 0.10.

Once we tuned the parameters of our model, we used random stratified split to segregate the data in the ratio of 3:1 training and testing. We scored the learned labels vs. the human labels for each of the 25 test samples. Consistent with our cross validation score, the results on the 25 objects show 80% accuracy. Refer Fig. 10.

Labeled ↑ \ Predicted →	*rc*	*rp*	*wc*	*wh*	*wp*	*wt*	
rc	3		1				4
rp		6		1			7
wc			1				1
wh		3					3
wp					3		3
wt						7	7
	3	9	2	1	3	7	25

Labeled ↑ \ Predicted →	*a.b*	*a.b.c*	*b*	*b.c*	*c*	
a.b	3					4
a.b.c		1	1			1
b			3		2	5
b.c			1	4	1	6
c					9	9
	3	1	5	4	12	25

Fig. 10. Grasp Type Confusion Matrix. Left: Predicted Grasps vs. Human Grasp Choices Right: Predicted Grasp Dimension vs. Human Grasp Dimensions

These accuracy scores need to be examined in the context of the discussion presented in section III-A. Since there are multiple grasp types feasible for each object, the true test of these learned models is to execute the predicted grasps on the physical robot.

C. Robot Grasp Execution and Validation

We then tested the grasp predictions on a set of ten objects unfamiliar/unknown to our learned model. We combined the grasp prediction and inverse kinematics modules and interfaced them with ROS Python libraries to control the AR10. Objects were placed in a convenient position and orientation to execute the grasps. We let the robot choose the grasp type using just the object description entered into Python console. If the grasp is secure during the lift/move maneuvers of the robotic arm, we label the grasp as success. We also had humans grasp the same objects and labeled the grasps for comparison.

The grasp type chosen by the learned model for 8 out of 10 matched human grasp types. There were two instances of failure on the robot. Of the one that failed (a smooth plastic cap), the grasp chosen by the model was same as human grasp, but the failure on AR10 could be attributed to inadequate friction. The other failed trial was on a wallet. The learned grasp choice on wallet was different from the human grasp label and the robot found it difficult to secure the object - a clear case of prediction error. An interesting observations was on the coffee can. The learned grasp on the coffee can was different from the human preferred grasp, yet the grasp was successful.

V. Discussion

Human hand has 20 joints and thousands of mechanoreceptors [11] with disproportionate amount of the brain's sensorimotor resources dedicated to grasping. It required multiple learning modules to replicate the basic human cognitive-motor grasping behavior. Based on the results, we were able to achieve reasonable success; helped in part by the use of humanoid robotic hand to grasp objects *designed* for human hands. The principal reason for confusion in predicting grasps is the decision to map unique grasp types to objects. A mislabeled grasp could very well be a feasible one, just not the one most preferred by a human. The example of the coffee can above, is evidence that we need to look at grasp choices as a probability distribution conditioned on the object features and not just as a unique label.

VI. Conclusion

This paper has presented an approach to parse object descriptions in natural language and determine the appropriate grasping strategy using the parsed object attributes. The framework of grasping strategy was developed using multiple machine learning models and performs reasonably well. Future studies can look at converting unique grasps labels to multi-label grasp probability distribution to better represent the human grasping behavior. The model can be scaled to include different task conditions and a more broader selection of objects. The neural network model could be easily extended to a deeper multi-layer network with additional features including object's position and orientation.

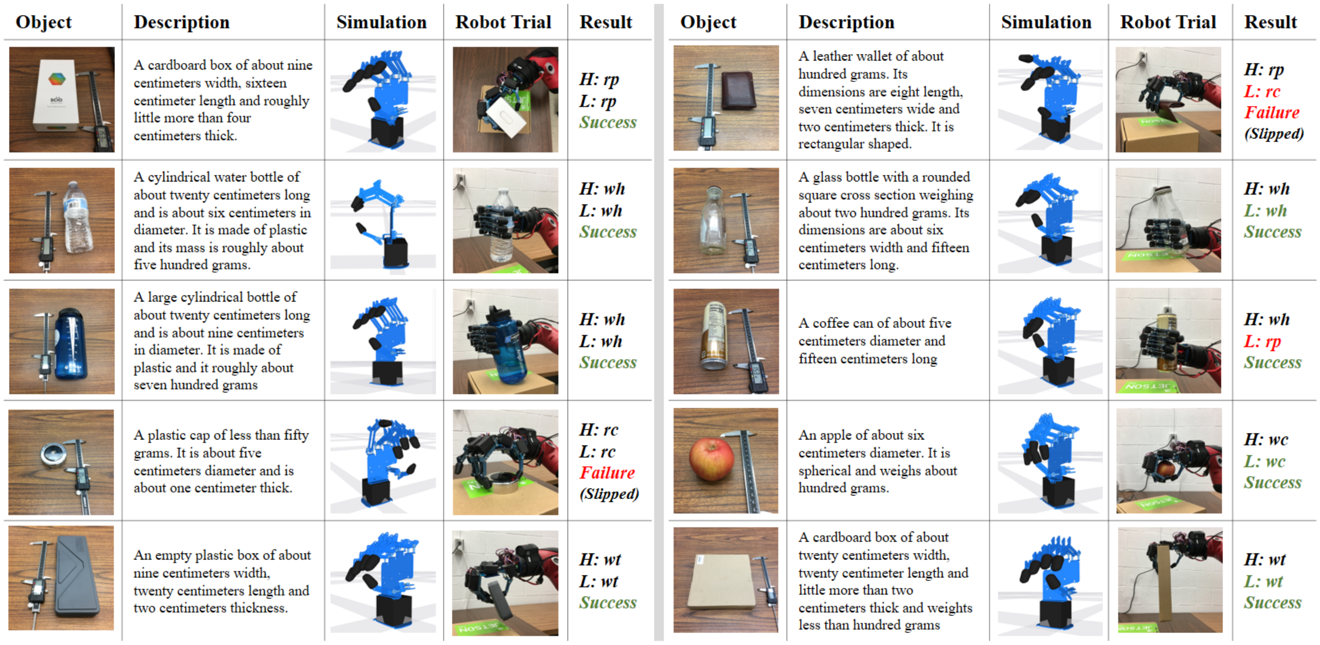
References

[1] P. Aivaliotis, A. Zampetis, G. Michalos, and S. Makris. A machine learning approach for visual recognition of complex parts in robotic manipulation. *Procedia Manufacturing*, 11:423 – 430, 2017. 27th International Conference on Flexible Automation and Intelligent Manufacturing, FAIM2017, 27-30 June 2017, Modena, Italy.

[2] J. Bao, Y. Jia, Y. Cheng, H. Tang, and N. Xi. Detecting target objects by natural language instructions using an rgb-d camera. *Sensors*, 16(12), 2016.

[3] P. Bezk, P. Bozek, and Y. Nikitin. Advanced robotic grasping system using deep learning. *Procedia Engineering*, 96, 12 2014.

[4] M. Bonilla, D. Resasco, M. Gabiccini, and A. Bicchi. Grasp planning with soft hands using bounding box object decomposition. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 518–523, Sept 2015.



Object	Description	Simulation	Robot Trial	Result
	A cardboard box of about nine centimeters width, sixteen centimeter length and roughly little more than four centimeters thick.			*H: rp* *L: rp* *Success*
	A cylindrical water bottle of about twenty centimeters long and is about six centimeters in diameter. It is made of plastic and its mass is roughly about five hundred grams.			*H: wh* *L: wh* *Success*
	A large cylindrical bottle of about twenty centimeters long and is about nine centimeters in diameter. It is made of plastic and it roughly about seven hundred grams			*H: wh* *L: wh* *Success*
	A plastic cap of less than fifty grams. It is about five centimeters diameter and is about one centimeter thick.			*H: rc* *L: rc* *Failure* *(Slipped)*
	An empty plastic box of about nine centimeters width, twenty centimeters length and two centimeters thickness.			*H: wt* *L: wt* *Success*
	A leather wallet of about hundred grams. Its dimensions are eight length, seven centimeters wide and two centimeters thick. It is rectangular shaped.			*H: rp* *L: rc* *Failure* *(Slipped)*
	A glass bottle with a rounded square cross section weighing about two hundred grams. Its dimensions are about six centimeters width and fifteen centimeters long.			*H: wh* *L: wh* *Success*
	A coffee can of about five centimeters diameter and fifteen centimeters long			*H: wh* *L: rp* *Success*
	An apple of about six centimeters diameter. It is spherical and weighs about hundred grams.			*H: wc* *L: wc* *Success*
	A cardboard box of about twenty centimeters width, twenty centimeter length and little more than two centimeters thick and weights less than hundred grams			*H: wt* *L: wt* *Success*

Fig. 11. Robot grasp trials with 10 test objects. H: stands for human labeled grasps; L: stands for Learned grasp. L in red color indicates mismatch with human grasp preference. 'Failure' indicates robot's inability to grasp the object.

[5] M. Ciocarlie, C. Goldfeder, and P. Allen. Dimensionality reduction for hand-independent dexterous robotic grasping. In *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3270–3275, Oct 2007.

[6] M. R. Cutkosky. On grasp choice, grasp models, and the design of hands for manufacturing tasks. *IEEE Transactions on Robotics and Automation*, 5(3):269–279, Jun 1989.

[7] A. Farhadi, I. Endres, D. Hoiem, and D. Forsyth. Describing objects by their attributes. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1778–1785, June 2009.

[8] T. Feix, J. Romero, H. B. Schmiedmayer, A. M. Dollar, and D. Kragic. The grasp taxonomy of human grasp types. *IEEE Transactions on Human-Machine Systems*, 46(1):66–77, Feb 2016.

[9] F. Heinemann, S. Puhlmann, C. Eppner, J. lvarez Ruiz, M. Maertens, and O. Brock. A taxonomy of human grasping behavior suitable for transfer to robotic hands. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4286–4291, May 2015.

[10] A. Herzog, P. Pastor, M. Kalakrishnan, L. Righetti, J. Bohg, T. Asfour, and S. Schaal. Learning of grasp selection based on shape-templates. *Autonomous Robots*, 36(1):51–65, Jan 2014.

[11] R. S. Johansson and J. R. Flanagan. Coding and use of tactile signals from the fingertips in object manipulation tasks. *Nature Reviews Neuroscience*, 10(5):345 – 359, 2009.

[12] O. Kroemer, R. Detry, J. Piater, and J. Peters. Combining active learning and reactive control for robot grasping. *Robotics and Autonomous Systems*, 58(9):1105 – 1116, 2010. Hybrid Control for Autonomous Systems.

[13] L. Y. Ku, E. G. Learned-Miller, and R. A. Grupen. Associating grasping with convolutional neural network features. *CoRR*, abs/1609.03947, 2016.

[14] I. Lenz, H. Lee, and A. Saxena. Deep learning for detecting robotic grasps. *The International Journal of Robotics Research*, 34(4-5):705–724, 2015.

[15] S. Levine, P. Pastor, A. Krizhevsky, and D. Quillen. Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection. *CoRR*, abs/1603.02199, 2016.

[16] M. Li, K. Hang, D. Kragic, and A. Billard. Dexterous grasping under shape uncertainty. *Robotics and Autonomous Systems*, 75:352 – 364, 2016.

[17] M. Madry, D. Song, and D. Kragic. From object categories to grasp transfer using probabilistic reasoning. In *2012 IEEE International Conference on Robotics and Automation*, pages 1716–1723, May 2012.

[18] 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. *CoRR*, abs/1703.09312, 2017.

[19] A. T. Miller, S. Knoop, H. I. Christensen, and P. K. Allen. Automatic grasp planning using shape primitives. In *2003 IEEE International Conference on Robotics and Automation (Cat. No.03CH37422)*, volume 2, pages 1824–1829 vol.2, Sept 2003.

[20] J. Napier. The prehensile movements of the human hand. *The Bone and Joint Journal*, 38 B(4), Nov 1956.

[21] D. Nyga, M. Picklum, and M. Beetz. What no robot has seen before. probabilistic interpretation of natural-language object descriptions. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4278–4285, May 2017.

[22] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[23] R. Pelossof, A. Miller, P. Allen, and T. Jebara. An svm learning approach to robotic grasping. *Proceedings - IEEE International Conference on Robotics and Automation*, 4:3512–3518, 01 2004.

[24] L. Pinto and A. Gupta. Supersizing self-supervision: Learning to grasp from 50k tries and 700 robot hours. *CoRR*, abs/1509.06825, 2015.

[25] M. Santello, M. Flanders, and J. Soechting. Postural hand synergies for tool use. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 18:10105–15, 12 1998.

[26] A. Saxena, J. Driemeyer, and A. Y. Ng. Robotic grasping of novel objects using vision. *The International Journal of Robotics Research*, 27(2):157–173, 2008.

[27] D. Song, K. Huebner, V. Kyrki, and D. Kragic. Learning task constraints for robot grasping using graphical models. In *in IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2010.

[28] A. J. Spiers, M. V. Liarokapis, B. Calli, and A. M. Dollar. Single-grasp object classification and feature extraction with simple robot hands and tactile sensors. *IEEE Transactions on Haptics*, 9(2):207–220, April 2016.

[29] S. Stansfield. Robotic grasping of unknown objects: A knowledge-based approach. *The International Journal of Robotics Research*, 10(4):314–326, 1991.

[30] J. Wang, K. Markert, and M. Everingham. Learning models for object recognition from natural language descriptions. In *Proceedings of the British Machine Vision Conference*, 2009.