BioGIGA: 6-DoF grasp detection via implicit representations with a human-like manipulator

Gabriel G. Iturralde Duenas
Computer Science Department
The University of Texas at Austin
Austin, Texas 78703
Email: giturral@utexas.edu

***Abstract*—Grasp detection in clutter requires the robot to reason about the 3D scene from incomplete and noisy perception. In this work, we draw insight that 3D reconstruction and grasp learning are two intimately connected tasks, both of which require a fine-grained understanding of local geometry details. GIGA has already implemented this sinergies between grasp affordance and 3D reconstruction through multi-task learning of a shared representation. It takes advantage of deep implicit functions, a continuous and memory-efficient representation, to enable differentiable training of both tasks. Our method is inspired on GIGA and adds a human like gripper design by mimicking some biomechanical and physiological grasping patterns of the human hand. We train the model on self-supervised grasp trials data in simulation. Evaluation is conducted on a clutter removal task, where the robot clears cluttered objects by grasping them one at a time. This is introduced in a simple, low cost and effective implementation that outperforms GIGA in over 5% in terms of grasp success rate and over 15% in previous state-of-the-art baseline algorithms.**

I. Introduction

Generating robust grasps from raw perception is an essential task for robots to physically interact with objects in unstructured environments. Reliable robotic grasping is challenging because there is high uncertainty in object properties such as object shape, pose, material properties, and mass [22]. One great remaining challenge in robot manipulation is achieving human dexterity on grasping [18]. Therefore, understanding the way human grasp objects, knowing the kinematic implications and limitations associated with each grasp, and knowing common used patterns is a key component on this goal [28]. For this task, the geometry and physical properties of objects from partially observed visual data, infer a proper grasp pose in 3D position and orientation, and move the gripper to the desired grasp configuration for execution. Here we consider the problem of 6-DoF grasp detection in clutter from 3D point cloud of the robot's on-board camera with a simple gripper design that is similar to acts based on human hand grasping patterns. Our goal is to predict the type of grasp pattern that the robot should used based on human hand biomechanics on a clutter of objects using GIGA [12] for partial point cloud.

Previous work on robot grasping has cast it as a geometry-centric task, typically assuming access to the 3D model of the object [12, 8]. Grasps are thereby generated thorough optimization of analytical models of constrains derived from geometry and dynamics. This problem makes it difficult to translate this algorithms from simulation to reality. One solution for that is to integrate a 3D reconstruction using deep learning that has translate the focus to a data-driven paradigm [1, 14]. Deep networks are trained using large-size datasets, either through manual labeling [11] or self exploration [14]. However, deep learning for grasping often suffers from limited generalization within the training domains.

GIGA [12] has combined the synergies between affordance of generalization and geometry, leading to a high accuracy on grasping on clutter and uncluttered set of objects in comparison with other state-of-the-art algorithms. One limitation associated with two finger grippers and optimal point of gripping is based on the lack of fine tuning capability while grasping that is given by the grasping pattern of the human thumbs and digit. The human grasp task can be classified into power, intermediate and precision grasping [7]. For power patterns there is a rigid relation between the object and the hand, and for precision handling the hand is able to perform intrinsic movements on the object. Taxonomy and biomechanics studies on human grasp types had determine that there is a vertical displacement of the thumb in the plane of the hand palm (this is related with power grips), and for precision grasping there is a side or horizontal displacement of the thumb in the plane of the hand palm [7].

GIGA applies a structured implicit neural representation for 6-DoF grasp detection. This method extracts structured grids from Truncated Signed Distance Function (TSDF) voxel grid fused from the input depth image. A local feature can be computed from the feature grids given a query 3D coordinate. This local feature is used by the implicit functions for estimating the grasp affordance (in the form of grasp quality, grasp orientation, and gripper width of a parallel jaw) and the 3D geometry (in the form of binary occupancy) at the query location. For this article, we use the grasping quality that was predicted during the training data of GIGA and use that score to determine the type of grasp pattern that we the gripper will adapt. If the grasping quality is between 0.5-1 a power pattern will be used, if it below 0.5 a precise pattern will be used to increase the success rate of cluttering and decluttering, because precise grasp is used to fine tune the grasping quality for objects that have more complex shapes.

II. RELATED WORK

A. Learning grasp detection

Pioneer studies on grasping have developed analytical methods based on the object models [6, 25].Grasp planning is intended to find a gripper configuration that maximizes a success (quality) metric [22]. In recent years, deep learning methods have gained increased attention for the grasping problem [**?** 26]. Dex-Net [22, 23] introduced a two-stage pipeline for top-down antipodal grasping. It first samples candidates 4-DoF grasps. The grasp quality of each grasp candidate is then assessed by a convolutional neural network. GPD [10] and PointPGD [21] tackled 6-DoF grasp detection in clutter with a 2 stage pipeline: VGN [2] predicts 6-DoF grasp in clutter with a one-stage from input depth-images. In most recent works, deep networks are trained with only grasp supervision. Furthermore, another line of works have focused on estimating the affordance of an object and then detect grasps based on estimated affordance [30, 19]. GIGA is trained through a structured neural representation jointly with self-supervised geometry and grasp supervisions [12]. Our model is an upgrade of GIGA that is trained following the same pipeline than GIGA but the we add another feature in the self-supervised portion that is biomechanical grasp patterns based on the existing grasp training that GIGA uses.

B. Geometry-aware grasping and 3D reconstruction

The strong connection between grasp detection and geometry reasoning has inspired a line of work on geometry-aware grasping. DGGN regularizes grasps through 3D geometry reconstructions by predicting a voxel occupancy grid from partial observations and evaluates grasp quality from feature of the reconstructed grid [33]. Furthermore, PointSDF learns 3D reconstruction via implicit functions and shares the learned geometry features with the grasp evaluation network [32]. On the contrary, GIGA improves grasp detection and focuses more on the 3D reconstruction of graspable regions via joint training of both tasks [12]. BioGIGA takes advantages of the already 3D reconstructed of graspable regions and determines the pattern of grasp depending on the grasp quality index.

C. Biomechanics of human grasping patterns

The complexity and variety of uses of the human hand makes the categorization and classification of hand function challenging. The hand has 15 joints that results in more than 20 DoFs [13]. Consequently, directly modeling hand shapes is difficult and involves specifying a large number of parameters [7]. The ”GRASP Taxonomy” project was funded by the European Union extracted 33 different grasp types according to their taxonomy. The taxonomy provides a common terminology to define human hand configurations and is important in many domains such as human–computer interaction in which an understanding of the human is on of the basis for a proper interface [7]. Previous works have classified each grasp based on the need the precision or power to be properly executed [27]. This idea has been further studied and a distinction between ”power grip” and ”precision handling” has been introduced [20]. In the power grip, there is a rigid relation between the object and the hand, which means that all movements of the object have to be evoked by the arm. For the precision handling, the hand is able to perform intrinsic movements on the object without having to move the arm [7]. A third category, the intermediate grasp combines elements of power and precision grasps that are present roughly in the same proportion [15]. For the scope of this paper just the power and precision grasps will be considered because of the nature of the cluttering and decluttering task. Figure 1 taken from [7] shows the three main human grasping patterns (including the intermediate grasp) and how the fingers are positioned for different kind of objects and grasp patterns.

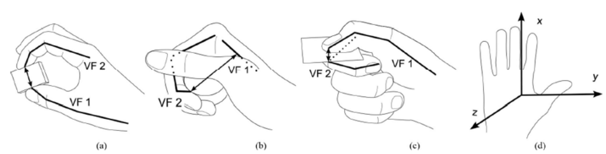


Fig. 1. Type of principal human grasp patterns with palm and fingers positioning for each pattern. [7]

Anatomically the power grasp is produced as a basic definition between hand surfaces along a direction generally parallel to the palm. This is associated with a vertical movement of the thumb with respect to the digits to produce higher power. Several robotic hand prosthesis have achieved this type of grasp by first placing the digits and having a delay on the move of the thumb [5], other approaches have approach that by providing higher force in the thumb while grasping the object [28] by using haptic sensors in the hand palm. For our design an increase and the thumb grasping force normal to the object point of contact and a delay on the thumb placement will be introduced by decreasing the velocity of the vertical movement of the thumb (one finger of the gripper) and its stiffness. Figure 2 taken from [7] shows our approach to mimic these two human grasp patterns.

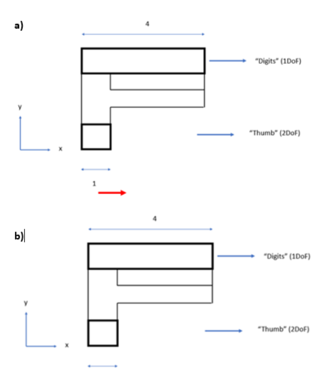


Fig. 2. Our approach using the BioGIGA gripper design based on human physiological behavior. a) Shows the precision grasp pattern in which there is an horizontal translation of the thumb gripper finger. b) Power grasp configuration in which the thumb is aligned with the index finger.

Furthermore, the human precision grasp is produced between hand surfaces along a direction transverse to the palm. There is an horizontal and vertical translation of the hand palm so as to place the thumb in between the index and middle fingers on a transversal plan. This type has been introduced on robotic hand prosthesis though a vertical and horizontal displacement of the thumb [28, 5]. Our design follows the same approach.

D. Implicit neural representations

Recent works have used the isosurface of an implicit function to represent the surface of a shape [4, 24]. By parametrizing these implicit functions with deep networks, they are capable of representing complex shapes smoothly and continuously in high resolution. The most common architecture for deep implicit functions is multi-layer perceptions (MLP), which encode the geometry information of the whole environment into the model parameters of the MLP [12]. However, they have difficulty in preserving the fine-grained geometric details of local regions. To solve this problem, hybrid representations have been introduced to combine feature grid structures and neural representations [29]. GIGA uses this hybrid approach for geometry reasoning in local objects parts [12]. Our model uses this same approach but redefines the output of this multilayer approach and classifies it into they type of grasp according to the likelyhood of achieving a successful grasp (grasp quality index).

III. METHODS

We introduce now BioGIGA, a learning algorithm inspired on GIGA [12] that exploits the synergy between affordance and geometry and 6-DoF grasp detection based on human physiological grasping patterns. We learn grasp affordance prediction and 3D occupancy prediction jointly with shared feature grids and a grasping pattern detection based on the grasp quality index that is one the outputs of the affordance prediction and a unified implicit neural representation. The grasp quality is defined as a scalar value between 0-1 that estimates the probability of grasp success. We learn to predict the grasp with binary success labels of executing the grasp trial in simulation with the training dataset. Figure 3 illustrates the overall model architecture.

A. Structure feature grids

To jointly learn the grasp affordance and 3D reconstruction, we need to extract a shared feature from the TSDF input. Following the same approach of GIGA, we adopt the encoder architecture from ConvONets [29] and learn to extract structure features grids from partial observation.

Our encoder takes as input a TSDF voxel field processes it with a 3D CNN layer to obtain a feature embedding for every voxel. Given these features, we construct planar feature representations by performing an orthographic projection onto a canonical plane for each input voxel. The canonical plane is discretized into pixel cells. Then we aggregate the features of voxels projected onto the same pixel cell using average pooling, which gives us a feature plane. The projection operation greatly reduces the computation cost while keeping the spatial distribution of feature points. We apply this feature projection and aggregation process to all three canonical frames. We therefore process each of these feature planes with a 2D U-Net which is composed of a series of downsampling and up-sampling convolutions with skip connections. The U-Net integrates both local and global information and acts as a feature inpainting network. The output feature grids denoted as c, are shared for affordance and geometry learning. This is exactly how GIGA architecture follows [12].

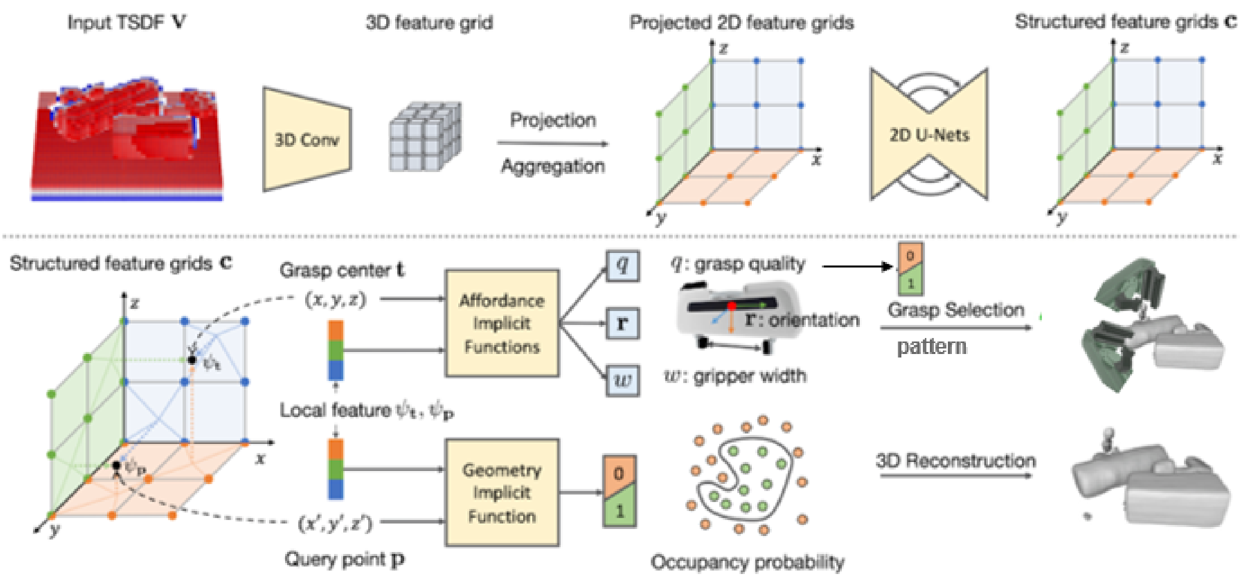
B. Implicit Neural Representations

1) Affordance Implicit Function: The affordance implicit function is represented by the grasp affordance field of grasp parameters (orientation and gripper width) and grasp quality. They map the grasp center t to grasp parameters of orientation and width (r and w) and the grasp quality metric q. These implicit neural representations enable learning directly from data with continuous grasp centers. In contrast, VGN has to snap grasp centers to the nearest voxel as it uses an explicit voxel-based grasp field. The snapping operation leads to information loss while our model does not. Furthermore, a threshold of 0.5 in the grasp quality metric was determined using GIGA training dataset on simulation for piles and clutters.

2) Geometry implicit function: Our geometry implicit function maps from an arbitrary query point inside the bounded volume to the occupancy probability at the point. Occupancy for purposes of this paper is defined as a binary value 0,1 indicating whether this point is occupied by any of the objects in the scene. This geometry implicit function is exactly defined as GIGA algorithm.

3) Grasp detection: GIGA takes as input a TSDF voxel grid, a grasp center, and multiple occupancy query points and predicts grasp parameters corresponding to the grasp center and occupancy probabilities at the query points. Given the trained GIGA model, we use a sampling procedure to select the final grasp pose. Grasp affordance is implicitly defined by the learned neural networks, so we need to query it from the learned implicit functions. To cover all possible graspable regions, we discretize the volume of the workspace into voxel grids and use the position of all the voxel cells as grasp centers. Then we query the grasp quality and grasp parameters corresponding to these grasp centers in parallel. Next, we mask out impractical grasps and apply nonmaxima suppression as done in VGN [2]. Finally, we select a grasp with the highest quality if the quality is beyond a threshold. If no grasp has the quality above the threshold, we do not make grasp predictions and give up the current scene. This works exactly as GIGA [12].

In the test dataset BioGIGA predicted a range of 0.5 and 0.75 that correspond to a precission pattern in order to have more fined-tuned grasp (similar to human biomechanics) and from 0.75 to 1 a power grasp pattern was predicted so as to have increase the accuracy of a successful grasp. Furthermore, the BioGIGA model was trained again using the same training dataset of GIGA for piles and clutters but using the new



[H]

Fig. 3. Model architecture of BIOGiga: This architecture is very similar to GIGA [12] but and additional parameter is added in the quality index that is the grasp pattern. Therefore, the affordance implicit functions predict grasp parameters and grasp pattern from the local feature at the grasp center. The geometry implicit function predicts occupancy probability from the local feature at the query point.

gripper configuration. That produced modified the ranges for the grasp quality index but the same threshold ranges were applied for selecting the type of grasping pattern. Figure **??** shows BioGIGA gripper design. For a precision grip the thumb gripper finger moves to the inside of the front of the palm and aligns in between the index and middle finger. For a power grasp configuration the thumb is aligned with the index finger but has a small time delay of 0.1-0.3 s and an increase in the gripper force that is introduced by modifying the velocity and stiffness of the thumb gripper.

Fig. 4. BioGIGA gripper design. For a precision grip the thumb gripper finger moves to the inside of the front of the palm and aligns in between the index and middle finger. For a power grasp configuration the thumb is aligned with the index finger but has a small time delay of 0.1-0.3 s and an increase in the gripper force that is introduced by modifying the velocity and stiffness of the thumb gripper.

IV. EXPERIMENTS

A. Experimental setup

Our model is trained in a self-supervised manner with ground-truth grasp labels collected from physical trials in simulation and occupancy data obtained from the object meshes. The use of TSDF enables zero-shot transfer of our model from simulation to a real Panda arm from Franka Emika. This is the exact similar environment as GIGA [12].

1) Simulation environment: Our simulated environment is built on PyBullet. We use a free gripper to sample grasps in a 30x30x30

$$cm^3$$

tabletop workspace. For a fair comparison, we use the same object assets as VGN, including 303 training and 40 test objects from different sources [16, 17, 3, 31]. The simulation grasp evaluations are all done with the test objects, which are excluded from training. We collect grasp data in a self-supervised fashion in two type of simulated scenes, pile and packed as in VGN. In the pile scenario, objects are randomly dropped to a box of the same size as the workspace. Removing the box leaves a cluttered pile of objects. In the packed scenario, a subset of taller objects is placed at random locations on the table at their canonical pose.

Once the scene is created a sample grasp centers and grasp orientations near the surface of the objects and execute these grasp samples in simulation. We store grasp parameters and the corresponding outcomes of the grasp trials and balance the dataset by discarding redundant negative samples. We collect the occupancy training data in the same scenes where grasp trials are performed. Upon the creation of a simulation scene, we query the binary occupancy of a large number of points uniformly distributed in the cubic workspace as the training data. This is the exact recreated simulation environment than GIGA [12].

2) Grasp execution: We select top grasps to execute by querying grasp parameters from the learned implicit functions with a set of grasp centers. For a fair comparison with VGN, our BioGIGA model samples 40x40x40 uniformly distributed grasp centers in the workspace and query the grasp parameters. However, our implicit representations are continuous, so we can query grasp samples in arbitrary resolutions. In BioGIGA (HR), we query at a higher resolution of 60x60x60. We use a set of clutter removal scenarios to evaluate BioGIGA and other baselines and the corresponfind GIGA, GIGA (HR) baselines. Each round, a pile or packed scene with 5 objects is generated. We take a depth image from the same viewpoint as training. The grasp detection algorithm generates a grasp proposal given the input TSDF. Until this point our algorithm is exactly similar to GIGA. [12]. Furthermore, given the orientation and grip width that result from this prediction a grasp pattern is selected according to the grasp quality metric. From a value of

$$q$$

from 0.5 to 0.75, a precision grasp is configured and from 0.75 to 1 a grasp power configuration is used. We execute the grasp and remove the grasped object from the workspace. If all objects are cleared, two consecutive failures happen, or no grasp is detected, we terminate the current scene. Otherwise, we collect the new observation and predict the next grasp. In our experiments, grasp proposals with a predicted grasp quality below 0.5 are discarded.

B. baselines

The baselines used for comparing our algorithm's performance where SHAF in which we used the highest point heuristic [9] by classic work of grasping in clutter, rather than the learned grasp quality, for grasp selection. In addition, GDP [10] (Grasp Pose Detection) that is a two stage 6-DoF grasp detection algorithm that generates a large set of grasp candidates and classifies each of them was set as the other baseline. VGN (Volumetric Grasping Network) a single-stage 6-DoF grasp detection algorithm that generates a large number of grasp parameters in parallel given input TSDF volume was also used. Moreover, GIGA-Aff [12] An ablated version of our method with only affordance implicit function branch. The network is trained with only grasp supervision but no reconstruction was also used. Lastly, both GIGA and GIGA(HD) were used as baseline comparison of BioGIGA.

Performance is measured using the following metrics averaged over 100 simulation rounds: 1) Grasp success rate (GSR), the ratio of success grasp executions; and 2) Declutter rate (DR), the average ratio of objects removed. The original VGN uses multi-view inputs, we re-train the VGN model on the same single-view data we used for training BioGIGA for fair comparisons. Those were the same criterion used by GIGA [12].

C. Grasp detection results

We report grasp success rate and declutter rate for different scenarios in Figure 5. It can be observed that BioGIGA and BioGIGA (HD) outperforms the baselines algorithm in almost all the categories, metrics and scenarios. The implicit neural representations learn to fit grasp affordance field withcontinuous functions. It allows us to query grasp parameters at a higher resolution as done in GIGA (HR) allow this algorithm to outperform the state-of-the-art algorithms by 10%. BioGIGA (HR) gives the highest performance in all cases and overperform GIGA (HR) by 5%.

This results were achieved using the same GIGA training dataset but using the new gripper configuration in the training dataset and in real simulation. Figure 6 shows the new gripper in the simulation environment performing a power grasping task because the quality index for this object was of 0.82. Initially, BioGIGA was training using the same gripper design as GIGA and then applying the new gripper configuration on the simulation environment. Here the grasping performance was around 3% higher than GIGA. Further analysis, showed that setting a grasp quality index threshold for no grasping decision of 0.45 and using BioGIGA configuration for the training fataset showed similar behavior in grasping metrics than GIGA.

Fig. 6. BioGIGA new gripper configuration for a grasping task using a power pattern in the simulation environment

V. Conclusions

The main contributions of this paper were: 1. Provide an effective and low-cost implementation for robotic gripper design that mimics some human grasping patterns that outperforms by 5% state-of-the-art grasping algorithms. 2. This algorithm applies the same concept that inspired GIGA related with the sinergies between affordance and geometry but outperforms GIGA in a 5% average and 15 % to baseline state-of-the-art algorithms. The main limitations of BioGIGA are based on that this method has not been proved in real world scenarios and that might have some challenging aspects relating to the gripper design, although the use of TSDF provides a smooth transition to real applications, and this new design strictly followed TSDF guidelines. Furthermore, our algorithm uses all the advantages and goals already achieved by GIGA that

Method	Packed		Pile	
	GSR (%)	DR (%)	GSR (%)	DR (%)
SHAF	56.6	58.0	50.7	42.6
GDP	35.4	30.7	17.7	9.2
VGN	75.4	79.2	60.7	44.0
GIGA-Aff	77.2	78.9	67.8	49.7
GIGA	83.5	84.3	69.3	49.8
GIGA (HR)	**87.9**	**86.0**	**69.8**	**51.1**
BioGIGA-Aff	85.4	86.9	75.3	53.4
BioGIGA	88.7	87.4	75.4	54.1
BioGIGA (HR)	**92.4**	**91.3**	**78.4**	**56.1**

Fig. 5. Quantitative results of clutter removal. We report mean and standard deviation of grasp success rates (GSR) and declutter rates (DR). HR denotes high resolution.

detects an optimal grasp through geometry and affordance functions. Lastly, there is a restriction related with the size of the gripper fingers that might be an issue for very small pick up object cluttering.

REFERENCES

[1] Jeannette Bohg, Antonio Morales, Tamim Asfour, and Danica Kragic. Data-driven grasp synthesis—a survey. *IEEE Transactions on Robotics*, 30(2):289–309, 2013.

[2] Michel Breyer, Jen Jen Chung, Lionel Ott, Roland Siegwart, and Juan Nieto. Volumetric grasping network: Real-time 6 dof grasp detection in clutter. *arXiv preprint arXiv:2101.01132*, 2021.

[3] Berk Calli, Arjun Singh, Aaron Walsman, Siddhartha Srinivasa, Pieter Abbeel, and Aaron M Dollar. The ycb object and model set: Towards common benchmarks for manipulation research. In *2015 international conference on advanced robotics (ICAR)*, pages 510–517. IEEE, 2015.

[4] Zhiqin Chen and Hao Zhang. Learning implicit fields for generative shape modeling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5939–5948, 2019.

[5] Joseph DeGol, Aadeel Akhtar, Bhargava Manja, and Timothy Bretl. Automatic grasp selection using a camera in a hand prosthesis. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 431–434. IEEE, 2016.

[6] Dan Ding, Yun-Hui Liu, and Shuguo Wang. Computing 3-d optimal form-closure grasps. In *Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No. 00CH37065)*, volume 4, pages 3573–3578. IEEE, 2000.

[7] Thomas Feix, Javier Romero, Heinz-Bodo Schmiedmayer, Aaron M Dollar, and Danica Kragic. The grasp taxonomy of human grasp types. *IEEE Transactions on human-machine systems*, 46(1):66–77, 2015.

[8] Carlo Ferrari and John F Canny. Planning optimal grasps. In *ICRA*, volume 3, page 6, 1992.

[9] David Fischinger, Markus Vincze, and Yun Jiang. Learning grasps for unknown objects in cluttered scenes. In *2013 IEEE international conference on robotics and automation*, pages 609–616. IEEE, 2013.

[10] Marcus Gualtieri, Andreas Ten Pas, Kate Saenko, and Robert Platt. High precision grasp pose detection in dense clutter. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 598–605. IEEE, 2016.

[11] Yun Jiang, Stephen Moseson, and Ashutosh Saxena. Efficient grasping from rgbd images: Learning using a new rectangle representation. In *2011 IEEE International conference on robotics and automation*, pages 3304–3311. IEEE, 2011.

[12] Zhenyu Jiang, Yifeng Zhu, Maxwell Svetlik, Kuan Fang, and Yuke Zhu. Synergies between affordance and geometry: 6-dof grasp detection via implicit representations. *arXiv preprint arXiv:2104.01542*, 2021.

[13] Lynette A Jones and Susan J Lederman. *Human hand function*. Oxford university press, 2006.

[14] Dmitry Kalashnikov, Alex Irpan, Peter Pastor, Julian Ibarz, Alexander Herzog, Eric Jang, Deirdre Quillen,

Ethan Holly, Mrinal Kalakrishnan, Vincent Vanhoucke, et al. Qt-opt: Scalable deep reinforcement learning for vision-based robotic manipulation. *arXiv preprint arXiv:1806.10293*, 2018.

[15] Noriko Kamakura, Michiko Matsuo, Harumi Ishii, Fumiko Mitsuboshi, and Yoriko Miura. Patterns of static prehension in normal hands. *American Journal of Occupational Therapy*, 34(7):437–445, 1980.

[16] Daniel Kappler, Jeannette Bohg, and Stefan Schaal. Leveraging big data for grasp planning. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4304–4311. IEEE, 2015.

[17] Alexander Kasper, Zhixing Xue, and Rüdiger Dillmann. The kit object models database: An object model database for object recognition, localization and manipulation in service robotics. *The International Journal of Robotics Research*, 31(8):927–934, 2012.

[18] Kilian Kleeberger, Richard Bormann, Werner Kraus, and Marco F Huber. A survey on learning-based robotic grasping. *Current Robotics Reports*, pages 1–11, 2020.

[19] Mia Kokic, Johannes A Stork, Joshua A Haustein, and Danica Kragic. Affordance detection for task-specific grasping using deep learning. In *2017 IEEE-RAS 17th International Conference on Humanoid Robotics (Humanoids)*, pages 91–98. IEEE, 2017.

[20] JMF1007266 Landsmeer. Power grip and precision handling. *Annals of the rheumatic diseases*, 21(2):164, 1962.

[21] Hongzhuo Liang, Xiaojian Ma, Shuang Li, Michael Görner, Song Tang, Bin Fang, Fuchun Sun, and Jianwei Zhang. Pointnetgpd: Detecting grasp configurations from point sets. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 3629–3635. IEEE, 2019.

[22] Jeffrey Mahler, Jacky Liang, Sherdil Niyaz, Michael Laskey, Richard Doan, Xinyu Liu, Juan Aparicio Ojea, and Ken Goldberg. Dex-net 2.0: Deep learning to plan robust grasps with synthetic point clouds and analytic grasp metrics. *arXiv preprint arXiv:1703.09312*, 2017.

[23] Jeffrey Mahler, Matthew Matl, Xinyu Liu, Albert Li, David Gealy, and Ken Goldberg. Dex-net 3.0: Computing robust vacuum suction grasp targets in point clouds using a new analytic model and deep learning. In *2018 IEEE International Conference on robotics and automation (ICRA)*, pages 5620–5627. IEEE, 2018.

[24] Lars Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. Occupancy networks: Learning 3d reconstruction in function space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4460–4470, 2019.

[25] Brian Mirtich and John Canny. Easily computable optimum grasps in 2-d and 3-d. In *Proceedings of the 1994 IEEE International Conference on Robotics and Automation*, pages 739–747. IEEE, 1994.

[26] Douglas Morrison, Peter Corke, and Jürgen Leitner. Closing the loop for robotic grasping: A real-time, generative grasp synthesis approach. *arXiv preprint arXiv:1804.05172*, 2018.

[27] John R Napier. The prehensile movements of the human hand. *The Journal of bone and joint surgery. British volume*, 38(4):902–913, 1956.

[28] Simon Ottenhaus, Daniel Renninghoff, Raphael Grimm, Fabio Ferreira, and Tamim Asfour. Visuo-haptic grasping of unknown objects based on gaussian process implicit surfaces and deep learning. In *2019 IEEE-RAS 19th International Conference on Humanoid Robots (Humanoids)*, pages 402–409. IEEE, 2019.

[29] Songyou Peng, Michael Niemeyer, Lars Mescheder, Marc Pollefeys, and Andreas Geiger. Convolutional occupancy networks. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*, pages 523–540. Springer, 2020.

[30] Christoph Pohl, Kevin Hitzler, Raphael Grimm, Antonio Zea, Uwe D Hanebeck, and Tamim Asfour. Affordance-based grasping and manipulation in real world applications. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 9569–9576. IEEE, 2020.

[31] Arjun Singh, James Sha, Karthik S Narayan, Tudor Achim, and Pieter Abbeel. Bigbird: A large-scale 3d database of object instances. In *2014 IEEE international conference on robotics and automation (ICRA)*, pages 509–516. IEEE, 2014.

[32] Mark Van der Merwe, Qingkai Lu, Balakumar Sundaralingam, Martin Matak, and Tucker Hermans. Learning continuous 3d reconstructions for geometrically aware grasping. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 11516–11522. IEEE, 2020.

[33] Xinchen Yan, Jasmined Hsu, Mohammad Khansari, Yunfei Bai, Arkanath Pathak, Abhinav Gupta, James Davidson, and Honglak Lee. Learning 6-dof grasping interaction via deep geometry-aware 3d representations. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3766–3773. IEEE, 2018.