ROBOTICS IN MANUFACTURING (JN PIRES, SECTION EDITOR)

A Survey on Learning-Based Robotic Grasping

Kilian Kleeberger [1] ● Richard Bormann [1] ● Werner Kraus [1] ● Marco F. Huber [1,2]

Abstract

Purpose of Review This review provides a comprehensive overview of machine learning approaches for vision-based robotic grasping and manipulation. Current trends and developments as well as various criteria for categorization of approaches are provided.

Recent Findings Model-free approaches are attractive due to their generalization capabilities to novel objects, but are mostly limited to top-down grasps and do not allow a precise object placement which can limit their applicability. In contrast, model-based methods allow a precise placement and aim for an automatic configuration without any human intervention to enable a fast and easy deployment.

Summary Both approaches to robotic grasping and manipulation with and without object-specific knowledge are discussed. Due to the large amount of data required to train AI-based approaches, simulations are an attractive choice for robot learning. This article also gives an overview of techniques and achievements in transfers from simulations to the real world.

Keywords Robotic grasping and manipulation · Artificial intelligence · Deep learning · Simulations · Sim-to-real transfer

Introduction

Humans see novel objects and can almost immediately determine how to pick them. The capabilities of robots lag far behind. Robotic grasping and manipulation is a critical challenge [1]. Creating cognitive robots that can operate at the same level of dexterity as humans has been approached for many decades. Despite the interest in research and industry, it remains an unsolved problem [2] [3].

This article is part of the Topical Collection on *Robotics in Manufacturing*

✉ Kilian Kleeberger
    kilian.kleeberger@ipa.fraunhofer.de

    Richard Bormann
    richard.bormann@ipa.fraunhofer.de

    Werner Kraus
    werner.kraus@ipa.fraunhofer.de

    Marco F. Huber
    marco.huber@ieee.org

[1] Fraunhofer IPA, Stuttgart, Germany

[2] IFF, University of Stuttgart, Stuttgart, Germany

Shorter product lifecycles and the steadily rising demand for customization require more flexible and changeable production systems leading to the need for an automatic configuration (Plug & Produce) of robot systems [4]. Developing robots that can operate in dynamic and unstructured environments (i.e., bin-picking, household or everyday environments, professional services) is of great interest. Approaches to robotic grasping utilize learning-based methods to automatically configure for the given task without any human intervention which allows to significantly reduce programming efforts [5]. Machine learning in particular is a promising approach to robotic grasping due to the generalization ability to novel objects.

This article aims to provide a comprehensive overview of different approaches to robotic grasping. A categorization of different methods is proposed as well as various techniques for grasping and sim-to-real transfer—motivated by the lack of real-world data—are introduced.

Categorization of Methods

Approaches to vision-based robotic grasping can be categorized along multiple different criteria. Generally speaking, approaches can be divided into analytic or data-driven methods

[6, 7]. Analytic (or sometime called geometric) approaches typically analyze the shape of a target object to identify a suitable grasp pose. Data-driven (or sometimes called empirical) approaches are based on machine learning and have gained popularity in recent years. They have made significant progress due to increased data availability, better computational resources, and algorithmic improvements. This review article focuses on learning-based approaches to robotic grasping and manipulation. For analytic grasping approaches, we refer the readers to [7–9].

Furthermore, approaches can be categorized as model-based or model-free, depending on whether or not specific knowledge about the object (e.g., CAD model or previously scanned model [10]) is used to solve the considered task. They can further be differentiated on whether they are focused on grasping and manipulating rigid, articulated, or flexible/deformable objects and whether the method is able to handle known, familiar, or unknown objects [6]. Figure 1 gives an overview of typical pipelines to robotic grasping. Model-based approaches for known rigid objects typically include a pose estimation step and allow a precise placement of the object. Model-free approaches directly propose grasp candidates and typically aim for a generalization to novel objects.

An additional criterion is the type of machine learning, i.e. whether the system is trained using supervised learning (SL) or reinforcement learning (RL) [11]. Annotations can be provided by humans or obtained in a self-supervised manner, i.e., the labels are generated automatically. Approaches typically either sample grasp candidates and rank them using a neural network (discriminative approaches) [12, 13] or directly generate suitable grasp poses (generative approaches) [14, 15]. Furthermore, approaches differ on whether they are trained in a simulation environment, in the real world, or both and utilize various kinds of sensor data (RGB image, depth image, RGB-D image, point cloud, potentially multiple sensors, …). Moreover, methods either operate in an open- (i.e., without any feedback) or closed-loop fashion [3, 16, 17]. Using continuous feedback based on visual features is commonly referred to as visual servoing [17]. Besides the robot hardware, the gripper type (two-finger gripper,

suction gripper, …) and gripper freedom (4D, 6D, …) also differentiate approaches. Moreover, some approaches focus on grasping of single separated objects only, while others target grasping in dense clutter. Furthermore, some methods are able to perform pre-grasp manipulations in order to move the object in a better configuration for grasping. Table 1 provides an overview of the discussed approaches and shows a small and exemplary selection from the variety of methods available in the literature. In addition to the abovementioned criteria, the reported grasp success rate is indicated, although being determined on different benchmarks.

Object Pose Estimation for Robotic Grasping

Model-based robotic grasping can be considered as a three-stage process where first object poses are estimated, then a grasp pose is determined, and finally a collision-free and kinematically feasible path is planned towards the object to pick it [34, 35]. This chapter focuses on the first part, which has the goal to estimate the translation and rotation relative to a given reference frame (usually the camera) of potentially multiple objects in the scene. This task is challenging because of sensor noise, varying lighting conditions, clutter and occlusions, and the variety of objects in the real world. Furthermore, object symmetries result in pose ambiguities which have to be addressed because with symmetries different annotations for identical observations are available [36–39]. For learning-based approaches on the second part, we refer the readers to [40].

When utilizing object-specific knowledge, approaches typically require an object-specific configuration (high amount of manual tuning) until a satisfactory system performance is reached which limits the scalability to novel objects [5]. More specifically, parameters for the template or feature matching method for pose estimation [41, 42] or the definition of robust grasp poses together with (static) priorities are required [35] and have to be tuned in real-world experiments. Therefore, model-based approaches aim for an automatic

Fig. 1 Typical pipelines to robotic grasping: Model-based approaches (top row) typically estimate the object pose, determine a suitable grasp pose on the object, plan a path, and finally execute the grasp. Model-free approaches (bottom row) directly determine grasp poses based on the observations given from the sensor. When being trained in simulation, sim-to-real techniques are needed for a robust transfer. This review article discusses the green elements of the figure
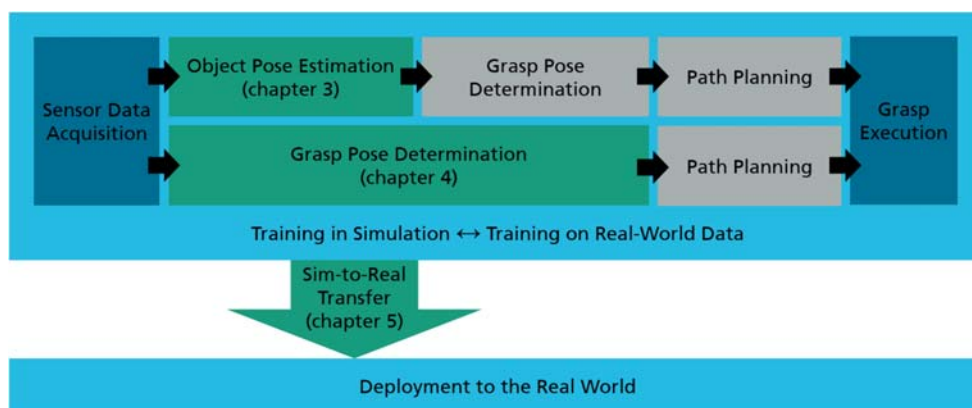
Table 1 Overview of the discussed approaches to robotic grasping and manipulation (selection)

	With object pose estimation?	Sensor data	Model-free or model-based?	For known, familiar, or unknown objects?	For rigid, articulated, or flexible/deformable objects?	Type of machine learning	Training in simulation or real world?	Open-loop or closed-loop	Discriminative or generative?	Gripper type	Gripper pose freedom	Precise object placement?	Grasping in clutter?	Reported grasp success rate (all on different benchmarks)	Pre-grasp manipulations?
DOPE [18]	Yes	RGB image	Model-based	Known	Rigid	SL	Simulation	Open-loop	–	–	6D	Yes	Yes	–	No
PPR-Net [19]	Yes	Point cloud	Model-based	Known	Rigid	SL	Simulation	Open-loop	–	–	6D	Yes	Yes	–	No
OP-Net [20•]	Yes	Depth image	Model-based	Known	Rigid	SL	Simulation	Open-loop	–	–	6D	Yes	Yes	–	No
SingleGrasp [14]	No	RGB-D image	Model-free	Unknown	Flexible	SL	Real world	–	Generative	Parallel-jaw	4D	No	No	–	No
MultiGrasp [14]	No	RGB-D image	Model-free	Unknown	Flexible	SL	Real world	–	Generative	Parallel-jaw	4D	No	No	–	No
Two Stage System [21]	No	RGB-D image	Model-free	Unknown	Flexible	SL	Real world	Open-loop	Discriminative	Parallel-jaw	4D	No	No	–	No
GG-CNN [15, 22]	No	Depth image	Model-free	Unknown	Flexible	SL	Real world/simulation	Closed-loop	Generative	Two finger	4D	No	Yes	–	No
Supersizing Self-supervision [23]	No	RGB images	Model-free	Unknown	Flexible	SL	Real world	Open-loop	Discriminative	Parallel-jaw	4D	No	Yes	–	No
Hand-Eye Coordination [24, 25•]	No	RGB image	Model-free	Unknown	Flexible	SL (RL)	Real world	Closed-loop	Discriminative	Two finger	4D	No	Yes	84–89%	No
Dex-Net [12, 13, 26, 27]	No	Depth image	Model-free	Unknown	Flexible	SL	Simulation	Closed-loop	Discriminative	Parallel jaw, suction	4D	No	Yes	81–88%	No
FC-GQ-CNN [28]	No	Depth image	Model-free	Unknown	Flexible	SL	Simulation	Open-loop	Discriminative	Parallel-jaw	4D	No	Yes	66%	No
QT-Opt [29••]	No	RGB image	Model-free	Unknown	Flexible	RL	Real world	Closed-loop	Discriminative	Two finger	4D	No	Yes	80–90%	No
QT-Opt RCAN [16]	No	RGB image	Model-free	Unknown	Flexible	RL	Simulation	Closed-loop	Discriminative	Parallel-jaw	4D	No	Yes	–	No
TossingBot [30]	No	RGB-D image	Model-free	Unknown	Flexible	SL	Simulation and real world	Open-loop	Discriminative	Two finger	4D	No	Yes	–	No
S4G [31]	No	Point cloud	Model-free	Unknown	Flexible	SL	Simulation	Open-loop	Discriminative	–	4D	No	Yes	–	No
GraspNet [32•]	No	Point cloud	Model-free	Unknown	Flexible	UL	Simulation	Open-loop	Generative	–	4D	No	Yes	–	No
Grasping in the Wild [33••]	No	RGB-D image	Model-free	Unknown	Flexible	RL	Real world	Closed-loop	Discriminative	Parallel jaw, suction	4D	No	Yes	Up to 98%	No

Table 1 (continued)

No	Generative	Parallel jaw	4D	No	Yes	85–87%
Yes	–	Two finger	4D	No	Yes	76–96%
Yes	–	Two finger	4D	No	Yes	70–94%
No	Discriminative	Two finger	4D	No	Yes	64–76%
No	Generative	Two finger	6D	No	Yes	77.1%
No	–	Parallel jaw	6D	No	No	88%
No	–	Two finger	6D	No	Yes	88–92%

configuration with minimal user input and without any tuning that has to be done by experts to allow a fast and easy transfer to novel objects.

Utilizing the strength of supervised learning for 6D object pose estimation requires large amounts of labeled data for training. Creating and annotating datasets with 6D poses is very tedious, time-consuming, and does not scale [43]. Thus, it is a trend to train models on synthetic data because simulations are an abundant source of data and flawless ground truth annotations are automatically available (see also "Simulations" section). Transfer techniques are used for deployment to the real world (see also "Techniques for Sim-to-Real Transfer" section). [18, 20•]

In recent years, research in 6D object pose estimation has been dominated by approaches based on convolutional neural networks (CNNs). Approaches typically either discretize the pose space in bins and predict a class [44, 45] or solve pose estimation in terms of a regression task [19, 20•, 46]. DOPE [18] uses a deep neural network to process an RGB image, outputs the 2D image coordinates of the 3D bounding box of the objects, and uses a PnP algorithm [47] to estimate the 6D pose of each instance. The model is trained entirely on synthetic data while for the transfer from simulation to the real world, DOPE employs a combination of domain randomization [48••] and photorealistic rendering. The authors further demonstrate that the pose estimator trained on synthetic data can operate in real-world grasping systems with sufficient accuracy.

Pose estimation challenges [49, 50] and standard benchmarking systems [51] for pose estimation allow advancing the state of the art and enable a transparent and fair comparison of different approaches. Especially, the robust pose estimation of multiple objects in bulk is a great challenge and of major importance. These scenarios, which are often present in industrial bin-picking scenarios, are challenging due to a high amount of clutter and occlusion as visualized

Fig. 2 Cluttered scene for bin-picking

in Fig. 2. A challenge focusing on 6D object pose estimation for bin-picking [49] has been organized at IROS 2019 and utilized a large-scale dataset [43] comprising fully 6D pose-annotated synthetic and real-world scenes. For evaluation, the metric from Brégier et al. [36, 37] was used which properly accounts for object symmetries and considers objects with visibility of more than 50%.

In general, learning-based approaches have proven to be robust to occlusions due to learning plausible object pose configurations [49]. PPR-Net [19], the winning method of the aforementioned challenge, operates on point clouds and utilizes PointNet++ [52] to estimate a 6D pose for each point of the point cloud and applies clustering in 6D space to compute the final pose hypotheses by averaging each identified cluster. The approach is outperformed by OP-Net [20•] in terms of average precision on the noisy Siléane dataset [36]. Furthermore, OP-Net is much faster than PPR-Net because it provides a much more compact parameterization of the output and does not require post processing. The approach discretizes the 3D space of the scene and regresses a pose and confidence for each resulting volume element.

A major advantage of learning-based object pose estimators is that they do not require a manual parameter tuning for the configuration of new objects [41, 42]. Furthermore, they can be entirely trained on synthetic data, which can easily be obtained using a physics simulation by dropping objects in a random position and orientation above a bin in the case of bin-picking [43] or by placing (household) objects in virtual scenes [18].

Model-Free Robotic Grasping

Model-free approaches are attractive due to their ability to generalize to unseen objects [53] and pose a dominant direction in robotic grasping research. They do not use prior knowledge about the objects and therefore work without a pose estimation step, which is in contrast to the approaches discussed in the "Object Pose Estimation for Robotic Grasping" section. Approaches often show promising results in terms of generalization ability to novel objects, and models are usually trained in an end-to-end fashion. A placement of the objects after picking is mainly not considered and the type of object being picked is unknown.

Supervised Learning for Robotic Grasping

Supervised learning is concerned with learning a (non-linear) mapping based on labeled training data. In this chapter, we categorize the approaches as discriminative or generative depending on whether the grasp configuration is the input or output.

Discriminative Approaches

Discriminative approaches sample grasp candidates (e.g., using CEM [54]) and rank them using a neural network. For grasp execution, the robot chooses the grasp with the highest score. These approaches typically have a high runtime because they require multiple forward passes of the neural network to get high-quality grasps. Nonetheless, these approaches come with the advantage that arbitrarily many grasp pose can be evaluated and these methods are not limited by discretization of the grasping primitives/output space. Furthermore, a gradient-based refinement process can be applied/employed to improve the grasp success rate [32•].

Levine et al. [24] proposed a learning-based approach to hand-eye coordination for robotic grasping based on RGB images. In their work, they used up to 14 robots to collect success labels for 800,000 grasps in 2 months. The trained convolutional neural network can predict the grasp success for a given candidate based on an RGB image of the bin and is used to servo the gripper towards successful grasps. While this approach demonstrates the potential of learning-based approaches to robotic grasping, changes in the hardware setup require the collection of new data for retraining the system.

Dex-Net [12, 26] uses a physics simulation to grasp objects in randomized poses on a plane. The outcome of the grasp is logged together with an aligned crop of a depth image where the grasp is located forming one sample to the dataset. Their Grasp Quality Convolutional Neural Network (GQ-CNN) is trained by using that dataset. The trained model can predict the grasp success for given grasp candidates and depth images and generalizes to different rigid, articulated, or flexible objects unseen during training. The Dex-Net framework has been extended to suction grippers [13] and a dual-arm robot [27] where the policy infers whether to use a parallel jaw or suction gripper for emptying a cluttered bin. Furthermore, a fully convolutional network architecture generating grasps has been proposed to avoid an expensive sampling and ranking of grasp candidates [28].

Generative Approaches

Generative approaches output a grasp configuration. One approach to this—called robotic grasp detection—is to detect oriented rectangles [55] in the image plane, which represent promising grasp candidates for parallel jaw grippers. This parameterization comprises the position, orientation, and opening width of the gripper as visualized in Fig. 3. The problem of robotic grasp detection is analogous to object detection [56–58] in computer vision with the only difference being an added term for the gripper orientation.

For the scenario where a single object is placed on a plane surface, Redmon et al. [14] proposed a system called SingleGrasp which can predict an oriented rectangle and
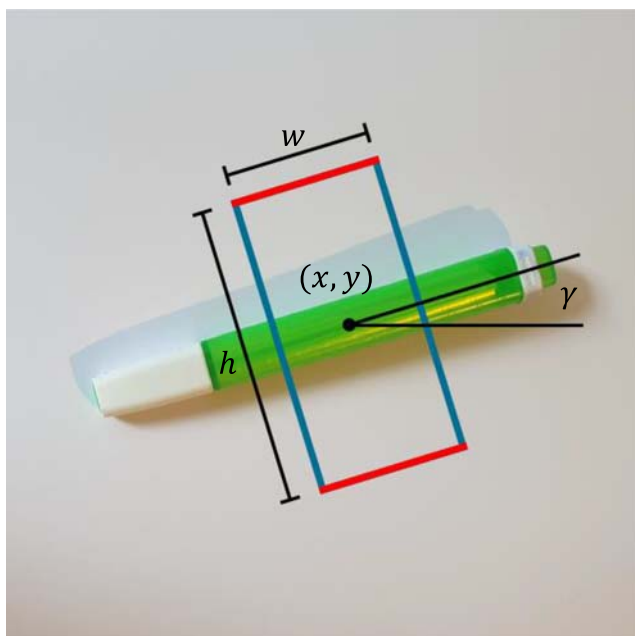
Fig. 3 Parameterization for robotic grasp detection: Two values for the position, two for the size, and one for the orientation of the oriented rectangle. Red sides indicate the jaws of the gripper and blue the opening width

simultaneously classify the object for a given RGB-D image using a neural network. Since an object can be grasped in multiple different ways, they also introduced MultiGrasp, which can predict multiple grasp poses per image. This approach led to the You Only Look Once (YOLO) [56, 57] approach for object detection. Lenz et al. [21] proposed a learning-based two-stage system that samples candidates and ranks them using a second neural network. In their work, they demonstrated that their approach can be used for real-world robotic grasping tasks. An increased performance is obtained by utilizing more sophisticated network architectures [3].

A public dataset for robotic grasp detection is the Cornell grasping dataset [59] which comprises 1035 images from 280 objects with human annotated grasps. Due to the low number of samples, the dataset has been heavily augmented for good performance [14]. The Jacquard dataset [60] comprises over 50,000 synthetic samples of more than 11,000 objects with grasps obtained from grasping trials in simulation and enables better generalization due to the increased diversity.

Utilizing these public datasets, GG-CNN [15, 22] outputs a grasp configuration together with a quality estimate for each pixel in the image using a small fully convolutional architecture. Due to its low computational demands, the approach can be used for closed-loop grasping in dynamic/non-static environments. Furthermore, this approach can grasp in clutter, although the model is trained on single isolated images only, which is due to the convolution being a local operation.

TossingBot [30] learns to throw arbitrary objects to given target locations which allows to increase the physical reachability of a robot arm. The authors propose an end-to-end formulation that jointly learns to infer control parameters for grasping and throwing from images of objects in a bin by trial and error. As a result, the system learns to select grasps that lead to predictable throws through self-supervision. The problem of throwing is simplified to predict the release velocity only. The release velocity is estimated using a physics-based controller and adjusted based on the residual estimate of the neural network.

Generative approaches are fast because they require one forward pass only. They usually provide multiple grasp candidates simultaneously and the highest quality grasp is executed by the robot.

Reinforcement Learning for Robotic Grasping and Manipulation

Deep reinforcement learning has emerged as a promising and powerful technique to automatically acquire control policies by trial and error. By processing raw sensory inputs, such as images, complex behaviors can be performed.

Pre-grasp manipulations such as pushing or shifting [61, 62] are also of major importance to rearrange cluttered objects and ensure that the objects can be grasped at all or more robustly. Using reinforcement learning, the trained policies also demonstrate generalization to novel objects [61, 62].

A comparison of a variety of methods based on deep reinforcement learning on grasping tasks is provided in [63]. QT-Opt [29••] demonstrates a rich set of manipulation strategies and responds dynamically to disturbances and perturbations. The robot observes a reward of 1 for successfully lifting an object and 0 for a failed grasp. Their closed-loop vision-based control framework operates in a similar setup as in [24, 25•, 64•] and reports a grasp success rate of 96% on unseen objects by optimizing long-horizon grasp success with a total of about 800 robot hours collected within 4 months and across 7 robots.

"Grasping in the Wild" [33••] allows a closed-loop 6D grasping of novel objects based on human demonstrations and can operate in dynamics scenes with moving objects, up to some speed constraint.

Simulations and Sim-to-Real Transfer

Despite all advantages w.r.t. performance and robustness, deep learning has the disadvantage of requiring large amounts of data for training. This is especially problematic in robotics, where the generation of training data on real-world systems can be expensive and time-consuming. For instance, Pinto et al. [23] trained a robot to grasp novel objects by collecting 50,000 trials in more than 700 h, Levine et al. [24, 25•] required 800,000 grasps parallelized over 14 robots in 2 months for robust grasping performance, and QT-Opt [29••] collected

over 560,000 grasps within the course of several weeks across 7 robots. Additionally, these systems are not invariant to changes in the hardware setup such as changing the gripper, table height, or moving the camera. To avoid the need to setup "arm farms" for learning robust robotic grasping and manipulation policies, using simulations is an attractive alternative.

Simulations

Commonly used physics simulations are V-REP/CoppeliaSim [65], PyRep [66], MoJuCo [67], Blender [68], and Gazebo [69], to name only a few. To overcome these aforementioned limitations, simulations can be employed because they provide an abundant source of data with flawless annotations. Furthermore, simulations are fast and can be parallelized across multiple machines for rapid learning or data generation. Physics simulations allow training the robots without wear and tear of the components and no interruption of production in the field. Apart from these advantages, simulations require the explicit programming of the desired application, potentially require license costs, and do not perfectly capture the properties of the real world.

Techniques for Sim-to-Real Transfer

Generally, models trained in simulations do not tend to directly transfer well to the real world due to the "reality gap" [64•, 70, 71]. This section discusses different approaches to allow bridging the simulation-to-reality gap. Models can be transferred to the real world by providing better simulations, domain randomization [48••], or domain adaptation [64•, 70, 72, 73].

Domain Randomization

The technique domain randomization [48••] applies various randomizations on the observations (vision randomization) or system dynamics (dynamics randomization) such that the real world appears to the model as just another variation. Randomizing various visual aspects of the simulator such as textures and colors of the objects and the background, lighting, object placement including camera placement, and type and amount of noise added to the image forces the network to learn to focus on the essential features of the image (vision randomization). Randomizations can also be applied to the dynamics of the system or environment [71] including gravity, mass of each link in the robot's body, damping of each joint, pose of the robot base as well as mass, friction, and damping of the manipulated objects (dynamics randomization) for a robust transfer from simulation to the real world.

This technique has been successfully used for object localization [48••], segmentation [74], robot control for pick-and-place [75], swing-peg-in-hole [76], opening a cabinet drawer [76], in-hand manipulation [77], one-handed Rubik's Cube

solving [78], precise 6D pose regression in highly cluttered environments [20•], etc. Modifications propose an automatic scheduling of the intensity of the randomization based on the current performance of the system [78] or adapting simulation randomizations by using real-world data to identify distributions that are particularly suited for a successful transfer [76]. Synthesizing millions of random object shapes for training [79] indicates further potentials of this technique for robotic grasping.

Domain Adaptation

Domain adaptation is a process that allows a machine learning model, trained with samples from a source domain to generalize to a target domain, which can be achieved by utilizing unlabeled data from the target domain. In sim-to-real transfer, the source domain is (usually) the simulation and the target domain is the real world. Prior work can be grouped into feature-level domain adaptation [80, 81], which focuses on learning domain-invariant features, and pixel-level domain adaptation [70], which focuses on restyling of images to bridge the domain gap [16, 64•].

Domain adaptation techniques are usually based on generative adversarial networks (GANs) [82]. With some unlabeled real-world data, those approaches allow a drastic reduction in the number of real-world samples needed. Using a similar system for hand-eye coordination as in [24, 25•], GraspGAN [64•] allows reducing the number of real-world samples needed to approximately 2% for similar system performance. This is a drastic reduction of the required real-world samples needed and allows a faster deployment of the solution in different setups.

Still, these approaches require data from the target domain (i.e., some samples from the real world are needed) which negatively affects scalability. Apart from being hard to train and often yielding fragile training results, the output images from the generator network (refiner) are not perfectly realistic and may include inaccuracies and artifacts.

RCAN [16] translates randomized simulation images to a canonical simulation version which are then used for policy training. The trained system can be used to translate real-world images to canonical images and consequently allows a sim-to-real transfer of the grasping policy, which is demonstrated by using QT-Opt [29••].

Benchmarking

As there are often many new approaches to pose estimation which are evaluated on a small number of datasets only, the Benchmark for 6D Object Pose Estimation BOP [51] aims for standardizing datasets to allow a better comparability. Apart from challenges such as "Occluded Object Challenge" [83],

SIXD [50], and "Object Pose Estimation Challenge for Bin-Picking" [49], BOP also organizes challenges for pose estimation.

Challenges focusing on robotic grasping and manipulation [84, 85] are of great value to the research community because of capturing and advancing the current state of the art in the field. The Amazon Picking/Robotics Challenge [2, 86–91] focused on autonomous picking in warehouse scenarios. Still, a participation can be challenging due to the required participation on site and hardware costs. Introducing detailed instructions on how to place the objects for picking [92] allows a comparison of different approaches. Especially, simulation environments allow a benchmarking of grasping and manipulation approaches under reproducible scenarios without hardware costs and are of high importance to measure scientific progress [93].

Conclusions

Learning-based approaches to robotic grasping enable picking of diverse sets of objects and are able to demonstrate high grasping success rates even in cluttered scenes and non-static environments. Machine learning and simulation allow fast and easy deployment due to the automatic configuration of model-based solutions and generalization abilities to novel objects of model-free approaches.

Despite impressive results, robotic grasping and manipulation is not solved. All discussed model-free approaches execute top-down grasps and have a limited flexibility in the gripper orientation. There is only a limited number of works focusing on learning-based approaches to robotic grasping in 6D for single objects [32•, 94–96] or in clutter [31, 33••, 34, 97]. While getting an increased focus in research, model-free grasping in 6D is especially relevant for picking objects from a cluttered bin [35], from a shelf [10], or for more robust grasps in general.

Usually, the task of the robot is to "grasp anything." Some works focus on a directed grasping to pick a specific object from a cluttered scene [63, 73, 98]. Model-free approaches do not allow a precise placement of the objects. Instead of simply dropping the picked object, many practical applications require an at least semi-precise or gentle placement of the components, which has been addressed less. While solutions for avoiding the entanglement of objects exist [99, 100], no general solution has been proposed to unhook complex object geometries.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by any of the authors.

References

Papers of particular interest, published recently, have been highlighted as:
• Of importance
•• Of major importance

1. Hodson R. A gripping problem: designing machines that can grasp and manipulate objects with anything approaching human levels of dexterity is first on the to-do list for robotics. In: Nature; 2018.
2. Zeng A, Song S, Yu K-T, Donlon E, Hogan FR, Bauza M, et al. Robotic pick-and-place of novel objects in clutter with multi-affordance grasping and cross-domain image matching. In: IEEE, editor. IEEE International Conference on Robotics and Automation (ICRA); May 21–25, 2018; Brisbane, QLD, Australia. PiscatawayJ: IEEE; 2018.
3. Kumra S, Kanan C. Robotic grasp detection using deep convolutional neural networks. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); September 24–28, 2017; Vancouver: IEEE; 2017.
4. Reinhart G, Hüttner S, Krug S. Automatic configuration of robot systems – upward and downward integration. In: Jeschke S, Liu H, Schilberg D, editors. Berlin, Heidelberg: Springer Berlin Heidelberg; 2011.
5. El-Shamouty M, Kleeberger K, Lämmle A, Huber M. Simulation-driven machine learning for robotics and automation. tm - Technisches Messen. 2019;86:673–84.
6. Bohg J, Morales A, Asfour T, Kragic D. Data-driven grasp synthesis—a survey. In: IEEE Transactions on Robotics (T-RO); 2014.
7. Sahbani A, El-Khoury S, Bidaud P. An overview of 3D object grasp synthesis algorithms. In: Robotics and Autonomous Systems; 2012.
8. Bicchi A, Kumar V. Robotic grasping and contact: a review. In: IEEE, editor. IEEE International Conference on Robotics and Automation (ICRA); April 24–28, 2000; San Francisco, CA, USA; 2000.

9. Shimoga KB. Robot grasp synthesis algorithms: a survey. In: The International Journal of Robotics Research (IJRR); 1996.

10. Bormann R, Brito BF de, Lindermayr J, Omainska M, Patel M. Towards automated order picking robots for warehouses and retail. In: Tzovaras, Dimitrios and Giakoumis, Dimitrios and Vincze, Markus and Argyros, Antonis, editor. Computer Vision Systems; September 23–25, 2019; Thessaloniki, Greece. Cham: Springer International Publishing; 2019.

11. Sutton RS, Barto AG. Reinforcement learning: an introduction. Cambridge Massachusetts: The MIT Press; 2018.

12. Mahler J, Liang J, Niyaz S, Laskey M, Doan R, Liu X, et al. Dex-Net 2.0: deep learning to plan robust grasps with synthetic point clouds and analytic grasp Metrics. In: Amato N, Srinivasa S, Ayanian N, Kuindersma S, editors. Robotics: Science and Systems (RSS); July 12–16, 2017; Cambridge, Massachusetts, USA: Robotics Science and Systems Foundation; 2017.

13. Mahler J, Matl M, Liu X, Li A, Gealy D, Goldberg K. Dex-Net 3.0: computing robust vacuum suction grasp targets in point clouds using a new analytic model and deep learning. In: IEEE, editor. IEEE International Conference on Robotics and Automation (ICRA); May 21–25, 2018; Brisbane, QLD, Australia. Piscataway, NJ: IEEE; 2018.

14. Redmon J, Angelova A. Real-time grasp detection using convolutional neural networks. In: IEEE, editor. IEEE International Conference on Robotics and Automation (ICRA); May 26–30, 2015; Seattle, WA, USA; 2015.

15. Morrison D, Leitner J, Corke P. Closing the loop for robotic grasping: a real-time, generative grasp synthesis approach. In: Kress-Gazit H, Srinivasa S, Atanasov N, editors. Robotics: Science and Systems (RSS); June 26–30, 2018. Pittsburgh: Robotics Science and Systems Foundation; 2018.

16. James S, Wohlhart P, Kalakrishnan M, Kalashnikov D, Irpan A, Ibarz J, et al. Sim-to-real via sim-to-sim: data-efficient robotic grasping via randomized-to-canonical adaptation networks. In: IEEE, editor. IEEE Conference on Computer Vision and Pattern Recognition (CVPR); June 16–20, 2019; Long Beach, CA; 2019.

17. Siciliano B, Khatib O, editors. Springer Handbook of Robotics. Berlin: Springer Science+Business Media; 2008.

18. Tremblay J, To T, Sundaralingam B, Xiang Y, Fox D, Birchfield S. Deep object pose estimation for semantic robotic grasping of household objects. In: Conference on Robot Learning (CoRL); October 29–31, 2018. Zürich: PMLR; 2018.

19. Dong Z, Liu S, Zhou T, Cheng H, Zeng L, Yu X, Liu H. PPR-Net: point-wise pose regression network for instance segmentation and 6D pose estimation in bin-picking scenarios. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); November 4–8, 2019; The Venetian Macao, Macau, China: IEEE; 2019.

20.• Kleeberger K, Huber MF. Single shot 6D object pose estimation. In: IEEE, editor. IEEE International Conference on Robotics and Automation (ICRA); May 31 – June 4, 2020; Palais des Congrès de Paris, France; 2020. **Provides state-of-the-art results for 6D object pose estimation in highly cluttered scenes.**

21. Lenz I, Lee H, Saxena A. Deep learning for detecting robotic grasps. In: The International Journal of Robotics Research (IJRR); 2015.

22. Morrison D, Corke P, Leitner J. Learning robust, real-time, reactive robotic grasping. In: The International Journal of Robotics Research (IJRR); 2019.

23. Pinto L, Gupta A. Supersizing self-supervision: learning to grasp from 50K tries and 700 robot hours. In: IEEE, editor. IEEE International Conference on Robotics and Automation (ICRA); May 16–21, 2016; Stockholm, Sweden; 2016.

24. Levine S, Pastor P, Krizhevsky A, Quillen D. Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection. In: International Symposium on Experimental Robotics (ISER); 2016.

25.• Levine S, Pastor P, Krizhevsky A, Ibarz J, Quillen D. Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection. In: The International Journal of Robotics Research (IJRR); 2018. **Highly influential work demonstrating the potential of deep learning for robotic grasping.**

26. Mahler J, Pokorny FT, Hou B, Roderick M, Laskey M, Aubry M, et al. 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: IEEE, editor. IEEE International Conference on Robotics and Automation (ICRA); May 16–21, 2016; Stockholm, Sweden; 2016.

27. Mahler J, Matl M, Satish V, Danielczuk M, DeRose B, McKinley S, Goldberg K. Learning ambidextrous robot grasping policies. SCIENCE ROBOTICS. 2019.

28. Satish V, Mahler J, Goldberg K. On-policy dataset synthesis for learning robot grasping policies using fully convolutional deep networks. In: IEEE Robotics and Automation Letters; 2019.

29.•• Kalashnikov D, Irpan A, Pastor P, Ibarz J, Herzog A, Jang E, et al. QT-Opt: scalable deep reinforcement learning for vision-based robotic manipulation. In: Conference on Robot Learning (CoRL); October 29–31, 2018; Zürich, Switzerland: PMLR; 2018. **Setting a milestone in robotic grasping and manipulation.**

30. Zeng A, Song S, Lee J, Rodriguez A, Funkhouser TA. TossingBot: learning to throw arbitrary objects with residual physics. In: Bicchi A, Kress-Gazit H, Hutchinson S, editors. Robotics: Science and Systems (RSS); June 22–26, 2019; Messe Freiburg, Germany; 2019.

31. Qin Y, Chen R, Zhu H, Song M, Xu J. S4G: amodal single-view single-shot SE(3) grasp detection in cluttered scenes. In: Conference on Robot Learning (CoRL); October 30 – November 1, 2019; Osaka, Japan; 2019.

32.• Mousavian A, Eppner C, Fox D. 6-DOF GraspNet: variational grasp generation for object manipulation. In: IEEE, editor. IEEE International Conference on Computer Vision (ICCV); October 27 – November 2, 2019; Seoul, Korea; 2019. **Addresses model-free grasping in 6D.**

33.•• Song S, Zeng A, Lee J, Funkhouser T. Grasping in the Wild: learning 6DoF closed-loop grasping from low-cost demonstrations. In: IEEE, editor. IEEE International Conference on Robotics and Automation (ICRA); May 31 – June 4, 2020; Palais des Congrès de Paris, France; 2020. **Addresses closed-loop model-free grasping in 6D and in cluttered scenes.**

34. ten Pas A, Gualtieri M, Saenko K, Platt R. Grasp pose detection in point clouds. In: The International Journal of Robotics Research (IJRR); 2017.

35. Spenrath F, Pott A. Gripping point determination for bin picking using heuristic search. In: CIRP Conference on Intelligent Computation in Manufacturing Engineering (CIRP ICME); July 20–22, 2016; Ischia, Italy; 2016.

36. Brégier R, Devernay F, Leyrit L, Crowley JL. Symmetry aware evaluation of 3D object detection and pose estimation in scenes of many parts in bulk. In: IEEE, editor. IEEE International Conference on Computer Vision (ICCV); October 22–29, 2017; Venice, Italy; 2017.

37. Brégier R, Devernay F, Leyrit L, Crowley JL. Defining the pose of any 3D rigid object and an associated distance. In: International Journal of Computer Vision (IJCV); 2018.

38. Hodaň T, Matas J, Obdržálek Š. On evaluation of 6D object pose estimation. In: European Conference on Computer Vision (ECCV); 2016.

39. Hinterstoisser S, Lepetit V, Ilic S, Holzer S, Bradski G, Konolige K, Navab N. Model based training, detection and pose estimation of texture-less 3D objects in heavily cluttered scenes. In: Asian Conference on Computer Vision (ACCV); 2012.

40. Spenrath F, Pott A. Using neural networks for heuristic grasp planning in random bin picking. In: IEEE International Conference on Automation Science and Engineering (CASE); August 20–24, 2018; Munich, Germany; 2018.

41. Ledermann T. Partikel-Schwarm-Optimierung zur Objektlageerkennung in Tiefendaten [Dissertation]. Stuttgart: University of Stuttgart; 2012.

42. Palzkill M. Heuristisches Suchverfahren zur Objektlageerkennung aus Punktewolken für industrielle Zuführsysteme [Dissertation]. Stuttgart: University of Stuttgart; 2014.

43. Kleeberger K, Landgraf C, Huber MF. Large-scale 6D object pose estimation dataset for industrial bin-picking. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); November 4–8, 2019. IEEE: The Venetian Macao, Macau, China; 2019.

44. Kehl W, Manhardt F, Tombari F, Ilic S, Navab N. SSD-6D: making RGB-based 3D detection and 6D pose estimation great again. In: IEEE, editor. IEEE International Conference on Computer Vision (ICCV); October 22–29, 2017; Venice, Italy; 2017.

45. Sundermeyer M, Marton Z, Durner M, Triebel R. Implicit 3D orientation learning for 6D object detection from RGB images. In: European Conference on Computer Vision (ECCV); 2018.

46. Tekin B, Sinha SN, Fua P. Real-time seamless single shot 6D object pose prediction. In: IEEE, editor. IEEE Conference on Computer Vision and Pattern Recognition (CVPR); June 18–22, 2018; Salt Lake City, Utah; 2018.

47. Lepetit V, Moreno-Noguer F, Fua P. EPnP: an accurate O(n) solution to the PnP problem. In: International Journal of Computer Vision (IJCV); 2009.

48.•• Tobin J, Fong R, Ray A, Schneider J, Zaremba W, Abbeel P. Domain randomization for transferring deep neural networks from simulation to the real world. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); September 24–28, 2017; Vancouver, BC, Canada: IEEE; 2017. **Highly influential work regarding sim-to-real transfer.**

49. Kleeberger K, Huber MF. Object pose estimation challenge for bin-picking. 2019. https://www.bin-picking.ai/en/competition.html. Accessed 1 June 2020.

50. Hodaň T, Michel F, Sahin C, Kim T-K, Matas J, Rother C. SIXD Challenge 2017. 2017. http://cmp.felk.cvut.cz/sixd/challenge2017/. Accessed 1 June 2020.

51. Hodaň T, Michel F, Brachmann E, Kehl W, Glent Buch A, Kraft D, et al. BOP: benchmark for 6D object pose estimation. In: European Conference on Computer Vision (ECCV); 2018. Accessed 1 June 2020.

52. Qi CR, Yi L, Su H, Guibas LJ. PointNet++: deep hierarchical feature learning on point sets in a metric space. In: I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett, editors. Advances in Neural Information Processing Systems 30 (NIPS 2017); December 04–09, 2017. Long Beach, California; 2017.

53. Saxena A, Driemeyer J, Ng AY. Robotic grasping of novel objects using vision. In: The International Journal of Robotics Research (IJRR); 2008.

54. Rubinstein RY, Kroese DP. The cross-entropy method: a unified approach to combinatorial optimization, Monte-Carlo Simulation and Machine Learning. Berlin: Springer-Verlag; 2004.

55. Jiang Y, Moseson S, Saxena A. Efficient grasping from RGBD images: learning using a new rectangle representation. In: IEEE, editor. IEEE International Conference on Robotics and Automation (ICRA); May 9–13, 2011; Shanghai, China. Piscataway, NJ: IEEE; 2011.

56. Redmon J, Divvala S, Girshick R, Farhadi A. You Only Look Once: unified, real-time object detection. In: IEEE, editor. IEEE Conference on Computer Vision and Pattern Recognition (CVPR); June 26 – July 1, 2016; Las Vegas, Nevada; 2016.

57. Redmon J, Farhadi A. YOLO9000: better, faster, stronger. In: IEEE, editor. IEEE Conference on Computer Vision and Pattern Recognition (CVPR); July 21–26, 2017; Honolulu, Hawaii; 2017. 7263–7271.

58. Szegedy C, Toshev A, Erhan D. Deep neural networks for object detection. In: C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, K. Q. Weinberger, editors. Advances in Neural Information Processing Systems 26 (NIPS 2013): Curran Associates, Inc; 2013.

59. Cornell University. Cornell Grasping Dataset. http://pr.cs.cornell.edu/grasping/rectdata/data.php. Accessed 1 June 2020.

60. Depierre A, Dellandréa E, Chen L. Jacquard: a large scale dataset for robotic grasp detection. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); October 1–5, 2018; Madrid, Spain: IEEE; 2018.

61. Zeng A, Song S, Welker S, Lee J, Rodriguez A, Funkhouser TA. Learning synergies between pushing and grasping with self-supervised deep reinforcement learning. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); October 1–5, 2018; Madrid, Spain: IEEE; 2018.

62. Berscheid L, Meißner P, Kroeger T. Robot learning of shifting objects for grasping in cluttered environments. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); November 4–8, 2019; The Venetian Macao, Macau, China: IEEE; 2019.

63. Quillen D, Jang E, Nachum O, Finn C, Ibarz J, Levine S. Deep reinforcement learning for vision-based robotic grasping: a simulated comparative evaluation of off-policy methods. In: IEEE, editor. IEEE International Conference on Robotics and Automation (ICRA); May 21–25, 2018; Brisbane, QLD, Australia. Piscataway, NJ: IEEE; 2018.

64.• Bousmalis K, Irpan A, Wohlhart P, Bai Y, Kelcey M, Kalakrishnan M, et al. Using simulation and domain adaptation to improve efficiency of deep robotic grasping. In: IEEE, editor. IEEE International Conference on Robotics and Automation (ICRA); May 21–25, 2018; Brisbane, QLD, Australia. Piscataway, NJ: IEEE; 2018. **Highly influential work regarding sim-to-real transfer for robotic grasping.**

65. Rohmer E, Singh SPN, Freese M. V-REP: a versatile and scalable robot simulation framework. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); November 3–7, 2013; Tokyo, Japan: IEEE; 2013.

66. James S, Freese M, Davison AJ. PyRep: bringing V-REP to deep robot learning; 26.06.2019.

67. Todorov E, Erez T, Tassa Y. MuJoCo: a physics engine for model-based control. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); October 7–12, 2012; Vilamoura, Algarve, Portugal: IEEE; 2012.

68. Blender. https://www.blender.org/. Accessed 1 June 2020.

69. Koenig N, Howard A. Design and use paradigms for gazebo, an open-source multi-robot simulator. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); 28 September – 2 October, 2004; Sendai, Japan: IEEE; 2004. p. 2149–2154. https://doi.org/10.1109/IROS.2004.1389727. Accessed 1 June 2020.

70. Bousmalis K, Silberman N, Dohan D, Erhan D, Krishnan D. Unsupervised pixel-level domain adaptation with generative adversarial networks. In: IEEE, editor. IEEE Conference on Computer Vision and Pattern Recognition (CVPR); July 21–26, 2017; Honolulu, Hawaii; 2017.

71. Peng XB, Andrychowicz M, Zaremba W, Abbeel P. Sim-to-real transfer of robotic control with dynamics randomization. In: IEEE, editor. IEEE International Conference on Robotics and Automation (ICRA); May 21–25, 2018; Brisbane, QLD, Australia. Piscataway, NJ: IEEE; 2018.

72. Shrivastava A, Pfister T, Tuzel O, Susskind J, Wang W, Webb R. Learning from simulated and unsupervised images through

adversarial training. In: IEEE, editor. IEEE International Conference on Computer Vision (ICCV); October 22–29, 2017; Venice, Italy; 2017.

73. Fang K, Bai Y, Hinterstoisser S, Savarese S, Kalakrishnan M. Multi-task domain adaptation for deep learning of instance grasping from simulation. In: IEEE, editor. IEEE International Conference on Robotics and Automation (ICRA); May 21–25, 2018; Brisbane, QLD, Australia. Piscataway, NJ: IEEE; 2018.

74. Danielczuk M, Matl M, Gupta S, Li A, Lee A, Mahler J, Goldberg K. Segmenting unknown 3D objects from real depth images using Mask R-CNN trained on synthetic data. In: IEEE, editor. IEEE International Conference on Robotics and Automation (ICRA); May 20–24, 2019; Montreal, Canada; 2019.

75. James S, Davison AJ, Johns E. Transferring end-to-end visuomotor control from simulation to real world for a multi-stage task. In: Conference on Robot Learning (CoRL); November 13–15, 2017; Mountain View, California: PMLR; 2017.

76. Chebotar Y, Handa A, Makoviychuk V, Macklin M, Issac J, Ratliff N, Fox D. Closing the sim-to-real loop: adapting simulation randomization with real world experience. In: IEEE, editor. IEEE International Conference on Robotics and Automation (ICRA); May 20–24, 2019; Montreal, Canada; 2019.

77. OpenAI, Andrychowicz M, Baker B, Chociej M, Jozefowicz R, Mc Grew B, et al. Learning dexterous in-hand manipulation; 01.08.2018.

78. OpenAI, Akkaya I, Andrychowicz M, Chociej M, Litwin M, McGrew B, et al. Solving Rubik's Cube with a robot hand; 16.10.2019.

79. Tobin J, Biewald L, Duan R, Andrychowicz M, Handa A, Kumar V, et al. Domain randomization and generative models for robotic grasping. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); October 1–5, 2018; Madrid, Spain: IEEE; 2018.

80. Ganin Y, Ustinova E, Ajakan H, Germain P, Larochelle H, Laviolette F, et al. Domain-adversarial training of neural networks. In: Journal of Machine Learning Research 17; 2016.

81. Bousmalis K, Trigeorgis G, Silberman N, Krishnan D, Erhan D. Domain separation networks. In: D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, R. Garnett, editors. Advances in Neural Information Processing Systems 29 (NIPS 2016): Curran Associates, Inc; 2016.

82. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. In: Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, K. Q. Weinberger, editors. Advances in Neural Information Processing Systems 27 (NIPS 2014); December 08–13, 2014. Palais des Congrès de Montréal, Montréal Canada: Curran Associates, Inc; 2014.

83. Visual Learning Lab Heidelberg. Occluded Object Challenge. 2015. https://hci.iwr.uni-heidelberg.de/vislearn/iccv2015-occlusion-challenge/. Accessed 1 June 2020.

84. Sun Y, Falco J, editors. Robotic grasping and manipulation: first robotic grasping and manipulation challenge, RGMC 2016, Held in Conjunction with IROS 2016, Daejeon, South Korea, October 10–12, 2016, Revised Papers. Cham: Springer; 2018.

85. Sun Y, Calli B, Falco J, Leitner J, Roa M, Xiong R, Yokokohji Y. Robotic grasping and manipulation competition. 2019. https://rpal.cse.usf.edu/competitioniros2019/. Accessed 1 June 2020.

86. Eppner C, Höfer S, Jonschkowski R, Martín-Martín R, Sieverling A, Wall V, Brock O. Lessons from the Amazon Picking Challenge: four aspects of building robotic systems. In: Hsu D, Amato N, Berman S, Jacobs S, editors. Robotics: Science and Systems (RSS); June 18–22, 2016; Ann Arbor, Michigan, USA; 2016.

87. Zeng A, Yu K-T, Song S, Suo D, Walker E, JR., Rodriguez A, Xiao J. Multi-view self-supervised deep learning for 6D pose estimation in the Amazon Picking Challenge. In: IEEE, editor. IEEE International Conference on Robotics and Automation (ICRA); May 29 – June 3, 2017; Singapore, Singapore: IEEE; 2017.

88. Morrison D, Tow AW, McTaggart M, Smith R, Kelly-Boxall N, Wade-McCue S, et al. Cartman: the low-cost cartesian manipulator that won the Amazon Robotics Challenge. In: IEEE, editor. IEEE International Conference on Robotics and Automation (ICRA); May 21–25, 2018; Brisbane, QLD, Australia. Piscataway, NJ: IEEE; 2018.

89. Hernandez C, Bharatheesha M, Ko W, Gaiser H, Tan J, van Deurzen K, et al. Team Delft's robot winner of the Amazon Picking Challenge 2016; 18.10.2016.

90. Jonschkowski R, Eppner C, Hofer S, Martin-Martin R, Brock O. Probabilistic multi-class segmentation for the Amazon Picking Challenge. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); October 9–14, 2016; Daejeon, South Korea: IEEE; 2016.

91. Correll N, Bekris KE, Berenson D, Brock O, Causo A, Hauser K, et al. Analysis and observations from the first Amazon Picking Challenge. In: IEEE Transactions on Automation Science and Engineering. p. 172–188.

92. Leitner J, Tow AW, Dean JE, Suenderhauf N, Durham JW, Cooper M, et al. The ACRV Picking Benchmark (APB): a robotic shelf picking benchmark to foster reproducible research. In: IEEE, editor. IEEE International Conference on Robotics and Automation (ICRA); May 29 – June 3, 2017; Singapore, Singapore: IEEE; 2017.

93. Ulbrich S, Kappler D, Asfour T, Vahrenkamp N, Bierbaum A, Przybylski M, Dillmann R. The OpenGRASP benchmarking suite: an environment for the comparative analysis of grasping and dexterous manipulation. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); September 25–30, 2011; San Francisco, CA, USA: IEEE; 2011.

94. Yan X, Hsu J, Khansari M, Bai Y, Pathak A, Gupta A, et al. Learning 6-DOF grasping interaction via deep geometry-aware 3D representations. In: IEEE, editor. IEEE International Conference on Robotics and Automation (ICRA); May 21–25, 2018; Brisbane, QLD, Australia. Piscataway, NJ: IEEE; 2018.

95. Zhou Y, Hauser K. 6DOF grasp planning by optimizing a deep learning scoring function. In: Amato N, Srinivasa S, Ayanian N, Kuindersma S, editors. Robotics: Science and Systems (RSS); July 12–16, 2017. Cambridge: Robotics Science and Systems Foundation; 2017.

96. Riedlinger MA, Völk M, Kleeberger K, Khalid MU, Bormann R. Model-free grasp learning framework based on physical simulation. In: International Symposium on Robotics (ISR). Munich, Germany; 2020.

97. Gualtieri M, Platt R. Learning 6-DoF grasping and pick-place using attention focus. In: Conference on Robot Learning (CoRL); October 29–31, 2018; Zürich, Switzerland: PMLR; 2018.

98. Jang E, Vijayanarasimhan S, Pastor P, Ibarz J, Levine S. End-to-end learning of semantic grasping. In: Conference on Robot Learning (CoRL); November 13–15, 2017; Mountain View, California: PMLR; 2017.

99. Matsumura R, Domae Y, Wan W, Harada K. Learning based robotic bin-picking for potentially tangled objects. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); November 4–8, 2019; The Venetian Macao, Macau, China: IEEE; 2019.

100. Moosmann M, Spenrath F, Kleeberger K, Khalid MU, Mönnig M, Rosport J, Bormann R. Increasing the robustness of random bin picking by avoiding grasps of entangled workpieces. In: CIRP Conference on Manufacturing Systems (CIRP CMS); July 1–3, 2020; Chicago, IL, US; 2020.