IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received September 14, 2020, accepted September 25, 2020, date of publication September 30, 2020, date of current version October 9, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3027923

Review of Deep Reinforcement Learning-Based Object Grasping: Techniques, Open Challenges, and Recommendations

MARWAN QAID MOHAMMED, KWEK LEE CHUNG, (Member, IEEE), AND CHUA SHING CHYI, (Senior Member, IEEE)
Faculty of Engineering and Technology, Multimedia University (MMU), Ayer Keroh 75450, Malaysia

Corresponding authors: Marwan Qaid Mohammed (marwan.qaid.mohammed@gmail.com) and Kwek Lee Chung (lckwek@mmu.edu.my)

This work was supported in part by the Multimedia University (MMU) through the MMU GRA Scheme under Grant MMUI/190004.02, and in part by the MMU Research and Development Capital Expenditure 2019 under Grant MMUI/CAPEX190008.

ABSTRACT The motivation behind our work is to review and analyze the most relevant studies on deep reinforcement learning-based object manipulation. Various studies are examined through a survey of existing literature and investigation of various aspects, namely, the intended applications, techniques applied, challenges faced by researchers and recommendations for minimizing obstacles. This review refers to all relevant articles on deep reinforcement learning-based object manipulation and solutions. The object grasping issue is a major manipulation challenge. Object grasping requires detection systems, methods and tools to facilitate efficient and fast agent training. Several studies have proposed that object grasping and its subtypes are the main elements in dealing with the environment and agent. Unlike other review articles, this review article provides different observations on deep reinforcement learning-based manipulation. The results of this comprehensive review of deep reinforcement learning in the manipulation field may be valuable for researchers and practitioners because they can expedite the establishment of important guidelines.

INDEX TERMS Deep reinforcement learning, object manipulation, robotic grasping.

I. INTRODUCTION

Grasping is an action of gripping and moving an object from one place to another. The three basic elements that must be considered during a grasping task are localization, object and environment; all of them require visual accuracy, robust sensing and fine control with consideration of slippage detection. Perception of the environment is one of the challenges that many researchers have addressed, and sensing is crucial to this task. The physical properties of robots and objects can be measured with sensors and transformed into signals that can be utilized by robot controllers. Sensors are essential for detecting actions in an environment and the way that a robot should move so that the behavior of the robot can be learnt as a result. By using sensors, a robotic system can be flexibly implemented in different workplaces to perform various tasks. The purpose of using sensors (e.g. vision and touch) is to help during the interaction between the robot's hand and an object within the robot's workspace. Global and local information can be acquired from sensors [1]. Global information is provided by vision sensors and used to determine the location of objects in the environment. The robot controller can exploit global information to avoid unwanted obstacles or move the end effector to its target successfully. Meanwhile, local information refers to the way the robot interacts with objects in the environment, and it is provided by touch sensors. Local information can be used by the robot controller to manipulate contacting objects or explore and extract the surface properties of objects [2]–[4].

Deep learning is used to train large artificial neural networks. Over the last decade, deep learning has elicited the attention of many researchers and led to advanced research on the application of robotics. Moreover, various deep learning algorithms have been developed and implemented in object manipulation/grasping. Hundreds of millions of parameters can be included in deep neural networks (DNNs) [5], [6].

The associate editor coordinating the review of this manuscript and approving it for publication was Huiyu Zhou.

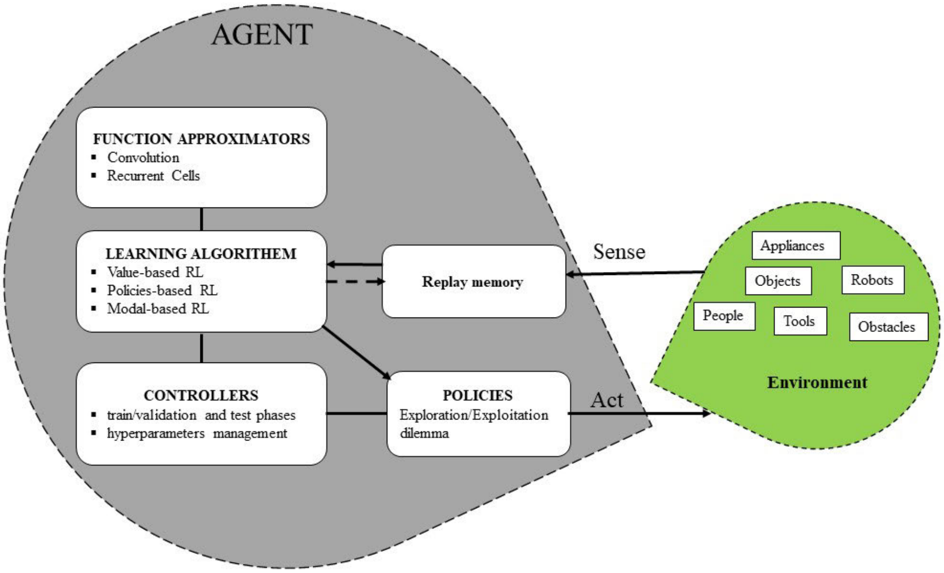


FIGURE 1. **Common framework of deep-RL methods [17].**

Deep learning continuously provides significant technological facilities for innovative applications and techniques. Specifically, extensive effort has been devoted to examine the use of deep learning (e.g. convolution neural networks (CNNs) [7]–[11] and deep belief neural networks (DBNNs) [12]–[14]) in developing sensor interpretation and control algorithms on the basis of actual data for particular challenges faced during grasping. Reinforcement learning (RL) plays an important role in robotics, particularly in object grasping. RL requires an agent to interact with the environment, and it is a way of learning the best policy via the trial-and-error strategy [15]. RL is an area of machine learning that focuses on how software agents should take actions in an environment, and it maximizes the notion of cumulative reward. Human-scale daily manipulation tasks require a reasoning component to make inferences on the basis of available knowledge [16]. Combining deep learning with RL creates a new approach called deep reinforcement learning (deep-RL). In standard deep-RL, the agent and the environment are the basis of learning. A general framework that extends beyond the agent and the environment and can exist in most deep-RL algorithms is illustrated in Fig.1.

In deep-RL, learning algorithms play a crucial role in improving the efficiency of robotic tasks, particularly grasping. Learning algorithms belong in one of three categories (value-based RL, policy-based RL and model-based RL), as illustrated in Fig. 2. Deep-RL approaches have been used in robotic manipulation [18], [19]; different applications, such as obtaining the best cooperation in manipulating objects [20] and learning the manipulation of deformable objects[21]; and manipulation tasks, such as reaching, grasping and placing [15], [22]–[25]. Moreover, deep-RL has produced promising results for soft robotic manipulators in comparison with hard robot manipulators [26], [27]. The recent technology wherein soft robotics are integrated with deep learning techniques has become pivotal in manipulation tasks. The rapid advancement in grasping capacity driven by vision-based complicated systems has motivated researchers to utilize models that artificially involve the use of intelligent robots. As a result, different techniques have been developed and used in the detection of robotic grasp [28] and delicate control for grasping objects in consideration of various sizes, shapes and structures [29]. These techniques have been improved by integrating DL and deep-RL algorithms into the robust structure of robotic grasping systems

Numerous reviews have been conducted in the robotic field. Several of them have examined deep learning methods involved in reaching and grasping tasks [30], [31]. RL was the focus in [32]. With the purpose of obtaining details on the cognitive capability for object surroundings during reaching and grasping, a thorough review was presented in [16], where recent and future works were also provided. However, this review did not cover deep-RL approaches. With regard to sensing, a review of visual and force/tactile control was conducted in [1] via a survey. Given the increasing demand for accurate and robust grasping, the tactile sensor has become a popular topic in robotics, and numerous studies have attempted to develop robotic sense of touch [3] or implement tactile sensing on robotic hands [33], [34]. With regard to the grasping mechanism design, several mechanisms were developed based on geckos and spiders by using dry adhesive materials [35], and others were created with elastic inflatable actuators [36]. Furthermore, soft robotic grippers were reviewed in [36]–[38]; these grippers were designed with advanced materials and soft components, such as silicone elastomers, active polymers and gels and shape

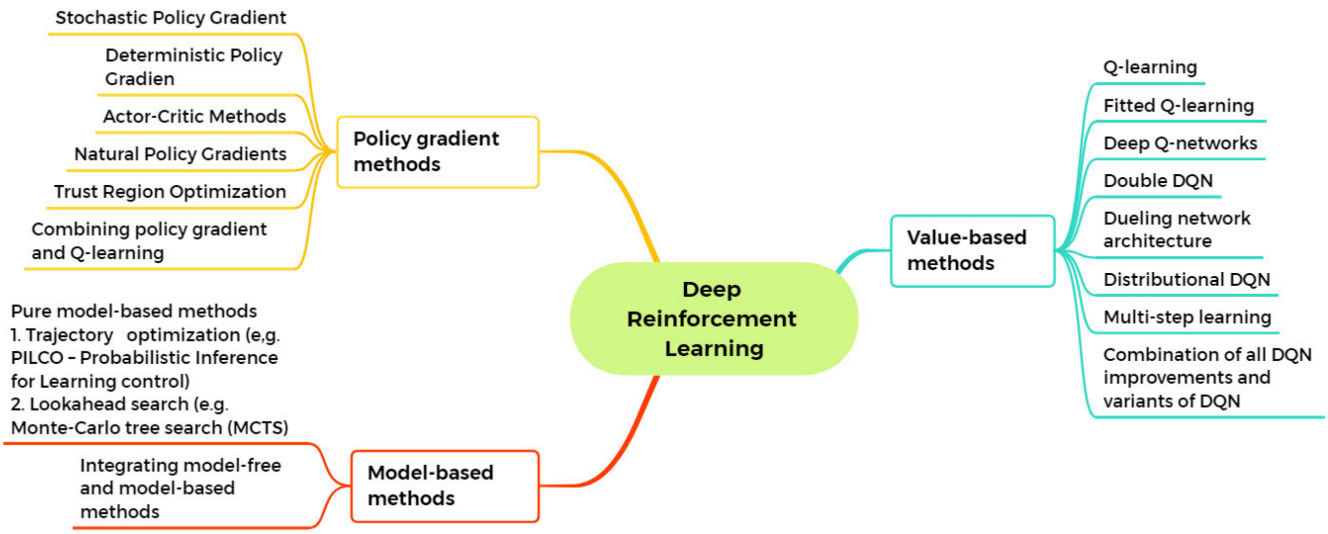


FIGURE 2. Learning algorithms-based reinforcement learning.

memory materials. These researchers focused on achieving light, simple and universal grippers by using the inherent functionality of materials. The most recent reviews on the application of deep-RL in robot manipulation are summarized in Table 1, which includes the year, title and description of the review papers.

Previous reviews have summarized different topics that are associated with robotics, as indicated in Table 1. The two most closely related topics to robotic manipulation were discussed in [50] and [52]. Ref. [50] highlighted the recent RL algorithms used in robot manipulation, but the range of topics was small due to the lack of papers pertaining to manipulation systems. Ref. [52] presented a review of studies that used machine learning for manipulation as a formalization of the robot manipulation learning problem. It attempted to build a bridge between such a computational algorithm concept and its application in the real world, which is not the focus in deep-RL associated with robotic grasping. In contrast to existing reviews that summarize the current status of robotics, the present review presents a comprehensive study of deep-RL-based object grasping and discusses the different aspects involved in robotics tasks by using the concept of deep-RL. We highlight the diversity of grasping learning problems that these approaches were implemented on and present future research directions and challenges. To the best of our knowledge, this is the first review that intensively studies the progress and new directions of deep RL-based robotic grasping. We expect this review to serve as an insightful reference for researchers in the robotic community.

This article begins by introducing the study and describing related review papers. Section II explains the review protocol. Section III describes grasping in clutter. Sim-to-real transfer is presented in Section IV, and learning from demonstration (LfD) and well-labelled data are highlighted in Sections V and VI, respectively. The next two sections present vision-based robotic grasping and other applications of deep-RL. The limitations of this work and future research directions are explained in Section IX. The last section concludes this work

II. REVIEW PROTOCOL

Review searches were conducted on four digital databases, namely:

- **Web of Science (WoS):** WoS is an indexing database that covers several academic disciplines.
- **ScienceDirect (SciDir):** SciDir provides access to science and technical journal articles.
- **IEEE Xplore (IEEEXplore):** IEEEXplore is a technical literature library in technology and engineering.
- **arXiv:** arXiv is a free distribution service and an open-access archive that covers several academic disciplines.

The selection of article was based on the index that formulates and facilitates both the simple and complex searches query, and specifically tracking several journals and conference articles on deep reinforcement learning based manipulation.

A. SELECTION OF STUDY

The procedure of the study selection involves intensive search of related articles that depends on two main iterations

- The titles and abstracts of the articles were scanned to exclude the duplication and unrelated articles.
- The full texts of the screened articles that included from the first iteration step were carefully read and the articles were organized in taxonomy.

B. SEARCH

Articles searching process was launched on 10th April 2020. The search query was specifically done on the WoS, SciDir, IEEEXplore, and arXiv databases using their search boxes running on full text with publication years between

TABLE 1. Review papers on closely related works.

Year	Work	Description of Work
2012	[39]	• Presented computational algorithms for generating 3D object grasps with autonomous multi-fingered robotic hands
2013	[40]	• Revealed links between physics and mathematics fields by reviewing works in reinforcement learning in terms of robot behaviour generation • Discussed the challenges and successes of applying reinforcement learning to robots
2016	[41]	• Focused on data-driven grasp synthesis and the methodologies for sampling and ranking candidate grasps
2017	[42]	• Focused on central algorithms in deep-RL (e.g. deep Q-network, trust region policy optimisation and asynchronous advantage actor–critic) • Highlighted the benefits of deep neural networks and provided a visual understanding of reinforcement learning
2018	[43]	• Reviewed and categorised deep-RL algorithms and their applications • Highlighted the pros and cons of algorithms • Pointed out the challenges that have been overcome by the DLR approach
2018 2018	[44] [17]	• Written as a teaching book and is more than a review article • Useful for those who wish to learn about deep learning, reinforcement learning, how they combine as deep-RL and their applications in the real world
2018	[45]	• Reviewed the basic algorithms of deep reinforcement learning in terms of research method, deep learning model involving RL and their application in various fields (e.g., AlphaGo, robotics and natural language processing) • General review of all reinforcement applications
2018	[46]	• Addressed the use of deep learning in region-based and region-free detection frameworks for object grasping
2018	[47]	• Presented the application of deep learning methods to generalised robotic grasp detection
2018	[48]	• Presented robotics-specific learning, reasoning and embodiment challenges in deep learning
2019	[49]	• Focused on the extreme end of the spectrum: how robots can acquire the learning capability through only a handful of trials and a few minutes; referred to this challenge as "micro-data reinforcement learning"
2019	[32]	• Reviewed deep reinforcement learning-based intelligent soft robotics, including various algorithms, and examples of their application in real-world scenarios
2019	[50]	• Highlighted the recent reinforcement learning algorithms used in robot manipulation • Small range of topics due to the lack of papers pertaining to manipulation systems
2019	[51]	• Presented an intensive study of vision-based robotic grasp detection methods
2019	[52]	• Presented a survey of studies that used machine learning for manipulation • Formalised the robot manipulation learning problem
2020	[53]	• Introduced a framework of personalisation settings and used it in a systematic literature review • Reviewed existing solutions and evaluated the corresponding strategies
2020	[54]	• Identified and discussed different aspects that influence a robotics task requiring the nontrivial use of the concept of affordances

2016 till 2020. For all mentioned databases, search was established by using the following keywords: **(("Robotic grasping" OR "Robotic manipulation" OR "Grasp∗" OR "Object manipulation" OR "Object Grasping" OR "Push") AND ("Deep Reinforcement Learning" OR "Deep Learning" OR "Reinforcement Learning"))**. The preferences of advanced search in all the search engines excluded books' chapters and other documents, to which only selected and relevant journals and conference articles written in English language were considered.

Therefore, in this review we will describe the current research topics found in the most recent literature, highlighting the most significant challenges. By searching in well-known digital database sources, such as WoS, SciDir, IEEEXplore, and arXiv where most of robotics journals are indexed, allows us to retrieve and review several hundreds of publications in the recent five years easily. We have restricted references to the most recent publications, mostly from 2016 to 2020. We aim to provide a solid knowledge of the recent research with an impression of the latest challenges that can be useful to other researchers in the future.

C. ELIGIBILITY CRITERIA

All the articles that met the criteria in Fig. 3 were included in the review. Mapping the space of the research on deep reinforcement learning based manipulation to a descriptive taxonomy comprising of four categories which are development scenario based on features analyzed, visualization, review

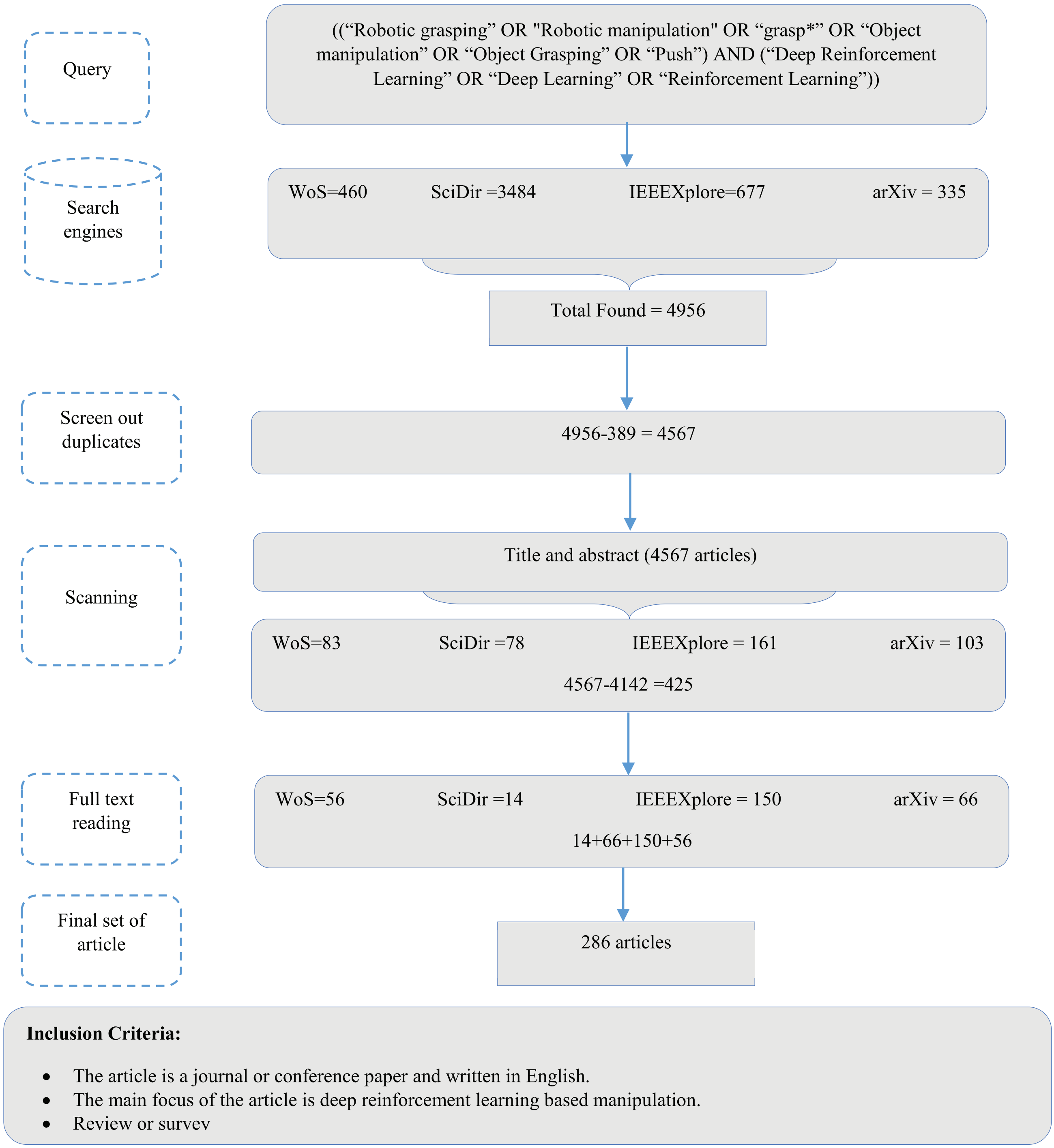


FIGURE 3. Selection of study, search query and inclusion criteria.

and survey, and evaluation and comparative study as an initial target was carried out. This categorization was derived from an intensive survey on the sources of literature. After the duplicated articles were removed, as shown in Fig. 3, all articles which did not meet the specified eligibility criteria were excluded. The exclusion criteria are defined as: (1) the non-English language articles. (2) The articles that discuss the grasping in general and do not focus on the deep reinforcement learning based manipulation.

III. GRASPING IN A CLUTTERED ENVIRONMENT

During grasping in a cluttered environment, a robot must be able to understand and recognize its surroundings and objects alike and perform a sequence of actions on the objects.

Although several studies have focused on object grasping or grasp pose detection in cluttered scenes, grasping an invisible object amongst other objects is challenging because robot arms need to explore the target position of that object. Most studies synergized nonprehensile primitive actions (e.g. push-to-grasp and shift-to-grasp) as a technique to facilitate object grasping in cluttered scenes. A robot with exploration capability based on nonprehensile primitive actions is the most common solution to create a free space around the target for performing grasping or placing tasks. This part focuses on grasping in a cluttered environment as one of the challenging robotic tasks due to the obstacles that prevent possible grasps. Grasping through a cluttered environment has been the focus of many researchers. According to our survey in this field, most existing studies synergized prehensile and nonprehensile primitive actions for complementarity. Here, nonprehensile (e.g. pushing) primitive action is one of the most common actions used with grasping to ease object grasping in clutter. Several studies utilized the suction grasp technique as an effective method without the need for synergy between prehensile and nonprehensile primitive actions. Others implemented a shifting action to create space for the gripper to put its fingers for grasping the target object. Grasping in cluttered scenes is performed based on four mechanisms of grasping, as illustrated in Fig. 4. To this end, grasping objects in clutter can be made easy once we introduce two actions that complement each other, which is also considered one of the most common solutions.

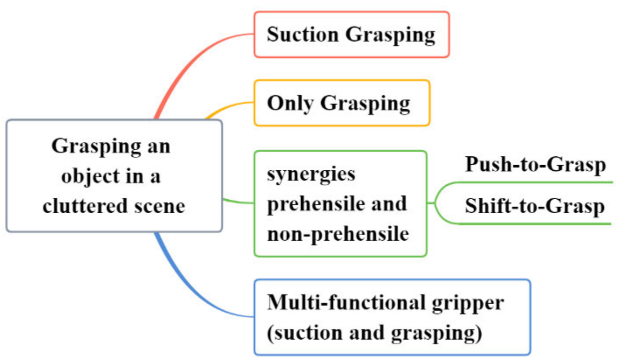


FIGURE 4. Mechanism of grasping in clutter.

A. GRASPING

Recently, deep-RL has been used in various robotic applications [40], such as placement [55], grasping objects mixed with towels [56], grasping deformable objects [57] and grasping in cluttered scenes [58], [59]. The grasping task in clutter has been intensively examined in numerous studies [60]–[63]. Deep-RL has led to advanced technologies by using visual and tactile features, particularly in robotic grasping [64]. It has also provided solutions for difficult tasks in clutter that are difficult to be automatically executed and repeated by using end-to-end training [65] based on trial and error [66], [67]. However, deep-RL-based robotic grasping, despite its merits, remains a challenge not only in building an interpretative model framework but also in terms of the complexity of the required resources. Robotic grasping has been categorized by extant studies into two categories. The first category covers analytic methods [68], which examine performance on the basis of modelling contact forces and the capability to resist external wrenches [69], [70] or constrain object movability during grasping [71]. The second category covers data-driven methods [41], which mainly depend on machine learning for developing models that map sensor readings to human labels [72] or physical trials [7]. This category explores the prospects of training model-agnostic deep grasping policies that detect grasps by exploiting learned visual features without explicitly using object-specific knowledge (i.e. shape, pose and dynamics). Pinto and Gupta [73] explored the use of pretrained models to improve the performance of deep policies during the execution of auxiliary tasks (e.g. poking). Most data-driven grasping algorithms at present can perform grasping detection in 6- degrees of freedom (6-DoF) with either closed-loop feedback, which only utilizes top–down grasps in simple tabletop settings [74], [75], or open-loop feedback [76], [77].

Many studies have concentrated on grasping or grasp pose detection, such as segmenting the target object [78], [79], combining red-green-blue color image with its corresponding depth image (i.e. RGB-D) based-multimodal data [80], template-based approach with convex hull [81], [82], use of a bounding box [83], [84] and detection of an object in a cluttered scene by using two-stream CNNs [85]. Grasping an unknown object in a cluttered scene has also been performed using CNN based on point cloud with a single depth camera [86], depth sensor [87] or 3D sensor [78], [88] under difficult environments due to the occlusion between the camera and objects. In [78], the grasping task was performed by learning a policy and fine-tuning grasp quality CNN (GQ-CNN) via Dex-Net. In addition, training Q-learning with pre-trained models (VGG) [89] that are fully convolutional neural networks (FCNs) has been implemented to perform grasping by capturing an RGB-D image from multiple views. Generative grasping CNN (GG-CNN) was presented in [90] to perform grasping in a cluttered environment.

Recent works used a single depth image, which is fed into a four-branch CNN (shared encoder–decoder), to perform grasping in clutter [91]. They utilized circumvent continuous motor control with direct mapping from pixels to Cartesian space inferred from the same depth image by using generative attention learning (GenerAL). Active vision has also been exploited to highlight the issue of object perception [92]–[94], which is a key subtask in targeted grasping. However, it is not commonly applied to grasping and manipulation problems in clutter [95], [96]. Instead of learning to pick and place objects by using planar manipulation (e.g. a single demonstrated goal state), Berscheid *et al.* [97] trained a robot to pick and place objects by using self-supervised learning without an object model. They combined robot learning of primitives estimated by FCNs and one-shot imitation learning. Furthermore, actor–critic deep-RL for visual grasping in clutter performs poorly in the grasping task, particularly the grasping of various types of objects from raw images with a sparse reward. Learning based on actor–critic deep-RL for visual grasping in clutter has been improved by using state representation learning (SRL) based on the disentanglement of a raw input image [98]. In clutter, focusing on the object axis configuration significantly increases the efficiency of grasping, but grasping from either the centroid of the object or along the major object axis is challenging due to complex-shaped objects. A real-time grasp pose estimation strategy was proposed in [99] for novel robotic pick-and-place applications. The approach estimates the object's contour in the point cloud and predicts the grasp pose and object skeleton on the image plane.

Most existing data-driven approaches attempt to perform grasping by using top–down planar grasping, which is inadequate for executing grasping in clutter. These approaches are insufficient for many real-application scenarios and greatly reduce grasp success. Several research groups considered this issue recently. For example, Qin *et al.* [100] addressed this issue in a challenging setting. They assumed that a group of household objects from unknown classes are irregularly scattered on a table. They proposed a framework for directly learning to retract 6-DoF grasps from the point cloud of the entire scene in one pass. In particular, they obtained a per-point scoring and a pose regression method for 6-DoF grasps. Real *et al.* [101] proposed a single view for performing grasping that closely works as the direct regression of point clouds. However, the back side of the object is unseen, leading to missed information. Grasping the back side would be difficult in this situation. Combining generative models with a new method of evaluating contact points appears to be the most effective way to increase the success rate. In [102], a method based on partial point cloud observations was implemented to plan 6-DoF grasps for the target object in a cluttered environment. The researchers mainly used an instance segmentation method to detect the target object. To generate a set of grasp points for the object, they adopted a cascaded approach by reasoning about grasps at an object level and then checking the cluttered environment for collisions. Zeng *et al.* [89] proposed a framework that segments and labels multiple views of a scene by using an FCN network. The approach leverages multi-view RGB-D data and self-supervised data-driven learning. Q-learning has also been trained based on FCN-DenseNet. However, these approaches based on template matching cannot perfectly deal with self-occlusion and mutual-occlusion between objects, which are common because of the inappropriate camera pose.

B. SUCTION AND MULTIFUNCTIONAL GRASPING

This section highlights the suction and multifunctional gripper as one of the mechanisms used to perform grasping in clutter. Suction grasping in clutter is a common and effective method with a high grasp success rate. For example, Zeng *et al.* [89] proposed a learning framework using FCNs that can efficiently model policies with affordances and improve run-times. Their framework uses multifunctional grasping (e.g. suction and grasping), which can improve the efficiency of grasping performance in clutter. However, their framework was designed based on RGB-D images with limited views, and they evaluated their framework with a limited type of objects. In another work, Real *et al.* [101] proposed a single view to perform grasping by using multifunctional grasping, which closely works as the direct regression of point clouds. Suction grasping can perfectly perform grasping in clutter with a higher success rate than multifunctional grasping. Shao *et al.* [103] performed grasping via suction grasping. Their proposed framework combines ResNet with the U-net structure, a special framework of CNN, to predict the picking region without recognition and pose estimation. Their framework was trained end-to-end with self-supervised learning. Although they presented a self-supervised approach to robotic bin picking that allows the robot to learn proper picking points in a cluttered bin, the authors only reported success in a simulated environment. Moreover, suction grasping is more effective in performing assembling tasks (e.g. kits) [67]. They leveraged data-driven shape previously learned from multiple kits during training. Their framework can easily be generalized to new objects and kits.

C. PREHENSILE AND NONPREHENSILE ACTIONS

Performing nonprehensile actions was one of the main problems in robotic manipulation in the past. The planning of nonprehensile motions, which emerged early by using classical solutions (including modelling dynamics of pushing with force friction [104], [105]), has developed rapidly. Many methods involved in the modelling field do not hold in practice [106], [107]. For instance, several factors (e.g. non-uniform friction distributions of object surfaces and variability of friction) can be prone to errors in predictions of friction-modelling pushing solutions in real settings.

To address the increasing demand for performing complex grasping tasks in clutter, Finn *et al.* [108] developed an action-conditioned video prediction system (Fig. 5) that can predict the distribution of pixel motions by using previous frames. The robot could generate nonprehensile actions (e.g. pushing objects) to push objects to desired locations, which helps the robot discover unseen objects during training. Although

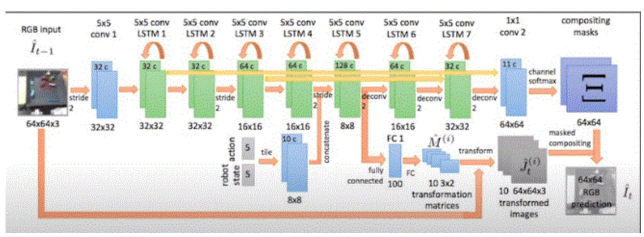


FIGURE 5. Set of normalised convolution filters that give rise to an independent Gaussian distribution over future images [108].

they utilized information from prior frames with model predictions without using the centroid point of the object, their approach requires large amounts of data to perform well in real-world situations. This is because they implemented unsupervised learning in their approach to physically interact through video prediction by dynamic neural advection (DNA). The authors collected a dataset of 59,000 pushing motions and applied it to different objects [109], as illustrated in Fig. 6. Although this method shows effective generalization to novel objects, it is constrained in terms of complex tasks and the time scales at which these tasks should be executed. In addition, this method cannot be applied to long-term planning and is only effective for short motions [110], [111]. To overcome this limitation, the authors updated their work in another paper by using a learning-based approach of hand–eye coordination for robotic grasping from monocular images [112]. They hope to achieve effective real-time control, which can be used to successfully grasp novel objects, and correct mistakes by continuous servoing. With regard to methodology, they trained a large CNN (Fig. 7) to predict the probability of grasps by using only monocular camera images and camera calibration or the current robot pose. However, these solutions cannot be generalized to new environments without being trained in the same environment, and they lack the capability to retain information on objects that are occluded during the predicted motion.

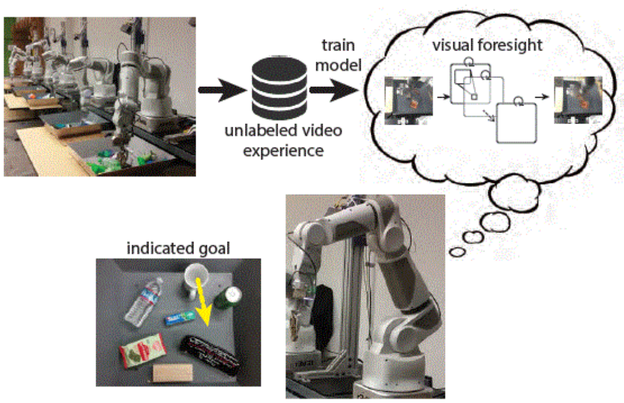


FIGURE 6. Video prediction model predicts stochastic pixel flow transformation from the current frame to the next frame [109].

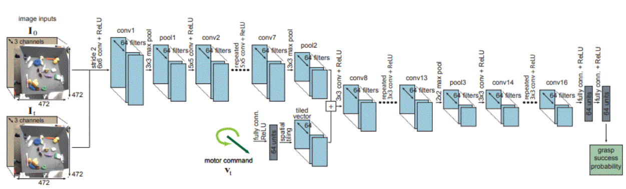


FIGURE 7. Convolutional dynamic neural advection (CDNA) that computes the expected value of motion distribution for every pixel [112].

In another work, Ebert et al. [113] used the DNA model, multilayer convolutional LSTM structure and skip connection neural advection model (SNA) (Fig. 8), which can be

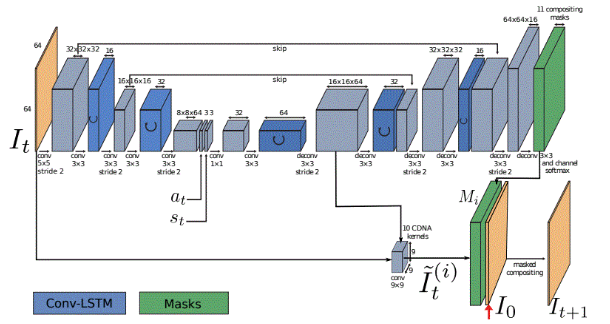


FIGURE 8. Spatial transformer predictors (STP) [113].

generalized to new environments without training in the same environment.

In terms of interleaving planning and execution in real time and closed-loop settings, Bejjani et al. [114] implemented a receding horizon planner (RHP) for pushing manipulation in clutter, as shown in Fig. 9. They addressed the problem of finding a suitable function-based heuristic for efficient planning and estimating the cost-to-go from the horizon to the goal. Their framework exploits the deep Q-learning (DQN) algorithm, which is trained using a dynamic neural network (DNN) to predict the actions to be executed. They also formulated an RL policy as RHP to select a random action with a probability of policy queries RHP for an action. However, the value function was learned over predefined features, which limits the framework's applicability to certain objects for a single particular shape. They overcame these limitations in their next work [115] by combining image-based learning systems with look-ahead planning. Other studies assumed object geometry [116] [117], and Cartesian coordinates were relied on to represent the state. Meanwhile, Song et al. [118] utilized object geometries to increase the efficiency of manipulation motions. Their approach addresses the problem where the goal is expressed in terms of numerous objects and all final poses. The algorithm computes solutions where the robot plows through a large collection of objects to separate and move different groups of objects apart. This manually designed function was replaced with a learned value function, similar to what has recently been done for the manipulation of movable obstacles (MAMO) problems. However, large-scale rearrangement planning (RP) problems remain unaddressed.

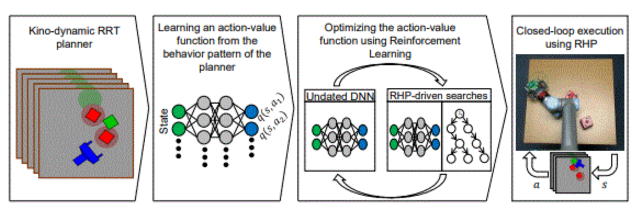


FIGURE 9. Receding horizon planner (RHP) for pushing manipulation in clutter [114].

Combining nonprehensile and prehensile manipulation policies plays a crucial role in grasping objects in clutter, but this area of research has not been intensively explored. Zeng *et al.* [119] presented a framework of grasping objects in cluttered scenes based on synergies pushing and grasping primitive actions. Their framework was built based on visual pushing–grasping (VPG), and Q-learning was trained with the DenseNet pre-training model, which is a fully connected network (DenseNet-FCN). However, the VPG framework was implemented to perform target-agnostic tasks and needed re-projection before inputting the prediction network. In another work, deep Q-critic was also trained on DenseNet-FCN to annotate the objects of interest and detect the existence of the target [120] by using a semantic segmentation module. Although the authors attempted to grasp an unseen object amongst other objects by pushing either the target or surrounding objects to create a free space around the target, no consideration was given to the materials of either the target or surrounding objects. Furthermore, the utilization of pushing primitive action as a method to singulate a target object from its surroundings in clutter was presented in [121]. The developed framework exploits the DQN algorithm to generate a new approach called split-DQN, which learns optimal push policies by trial and error. However, the framework is limited to objects with a certain shape, size and texture that are used in training and testing. Hundt *et al.* [122] described the schedule for positive task (SPOT) reward and the SPOT-Q RL algorithm, which efficiently learn multi-step block manipulation tasks in simulation and real-world environments. However, their framework focuses on how to stack objects on each other with combining push and grasp actions. In [123], shifting objects was determined to be another mechanism for facilitating object grasping in clutter by placing a finger on the top of the target object in such a way that the grasp probability increases. In this work, CNN with Q-learning was trained using argmax Q(s; a). By contrast, shifting objects by using a gripper may be reversely affected by the friction between the object and ground due to the type of material of the workspace and the object. Thus, shift action can be useful on certain grounds, but it can be ineffective on other grounds. This matter can be a future work direction for researchers.

RL-based, training-optimal push policies were presented in [124] for given depth observations of a scene. These policies facilitate grasping objects in cluttered scenes where the target can be invisible by using a deep neural network algorithm with Q-learning. However, the authors employed an instance push policy, in which a sole push policy is learned via Q-learning for the visible target in clutter. This approach can be improved further by using a more complex action repertoire, such as pushing in numerous directions and heights and picking up and removing objects. Residual policy learning (RPL) is a simple method that can improve non-differentiable policies by using model-free deep-RL to learn a parametric policy for vision-based manipulation. In [125], the RPL concept was implemented to perform the task of pushing, picking and placing objects in a cluttered scene by using deep deterministic policy gradients (DDPG) and an actor-critic algorithm, which works well in domains of continuous states and actions. In DDPG, the actor is updated following the deterministic policy gradient. However, the initial policy is an expert or feedback controller rather than a generative model.

D. SUMMARY

Table 2 summarizes related studies on grasping objects in clutter and presents useful information, including deep-RL methods, actions performed, frameworks and codes.

Grasping of objects, which are aligned to the edge or corner of the totebox, or stacked in a pile, remains a challenging task for a robot. In [126], using deep RL for combining grasping with pushing action has been presented to alleviate that issue with exploiting double experience replay. Another work has proposed framework based on a model-free Deep Reinforcement learning [127] to train control policies for exploiting visual information and proprioceptive states of the robot, in order to autonomously discover the robustness of pregrasp manipulation. Practically, the robot arm was trained to execute a sequential action, starting by pushing the object towards a support surface, and then establishing a pivot to lift up one side of the object, so that it could create a clearance between the object and the table as possible grasping solutions. In addition, a depth difference image-based bin-picking (DBP) algorithm has been proposed in [128], which does not need a neural network, because DBP has the ability of prediction the grasp pose from the object and its surroundings, which are obtained through depth filtering and clustering. In this approach, the object region was estimated by the density-based spatial clustering of applications with noise (DBSCAN) algorithm, and a depth difference image (DDI) that represents the depth difference between adjacent areas is defined. Different frameworks have been presented in achieving grasping object in clutter such as active affordance exploration framework which leverages the privileges of affordance map and the active exploration [129], integrating perception, action selection, and manipulation policies to address a version of the Mechanical Search problem [130], actor model with neural network that combines Gaussian mixture and normalizing flows [131], joint learning of instance and semantic segmentation for robotic pick-and-place with heavy occlusions in clutter [132], and predicting the quality and the pose of grasp using U-Grasping fully convolutional neural network(UG-Net) based on pixel-wise using depth image [133].

Moreover, Guo *et al.* [134] learned push skills based on combining You Only Look Once (YOLO) as detection algorithm with deep deterministic policy gradient (DDPG) algorithms. Han *et al.* [135] trained Q-Learning to learn suction grasp in clutter based on two CNN a fast region estimation network (FRE-Net) to predict which region contains pickable objects, and a suction grasp point affordance network (SGPA-Net) to determine which point in that region is pickable. Whereas, the issues of robotic manipulation for multiple object in clutter has been addressed in [136],

TABLE 2. Summary of works and their useful links.

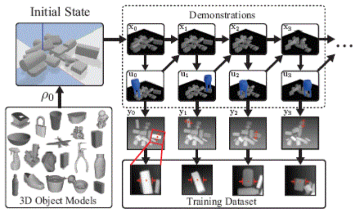
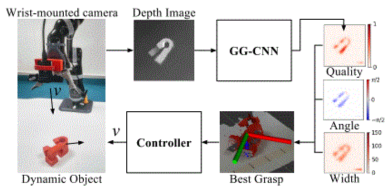
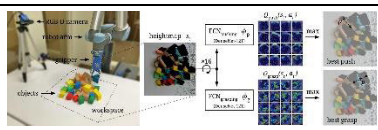
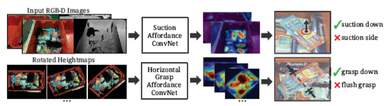
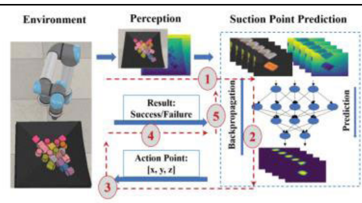
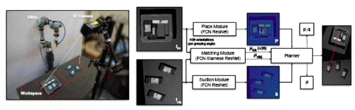
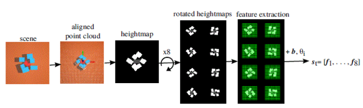
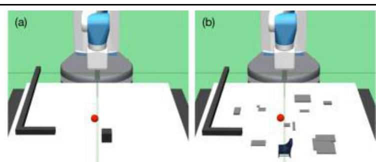
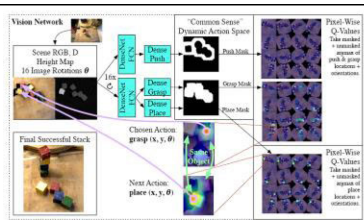
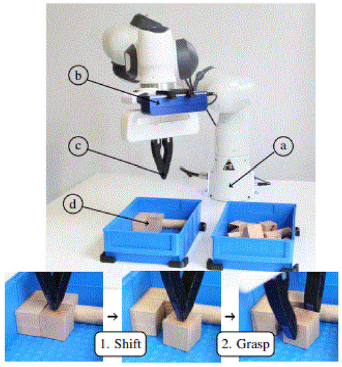
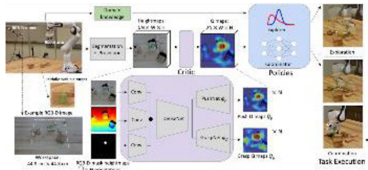
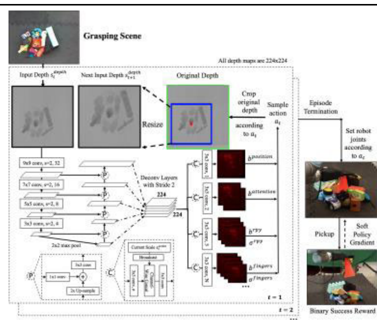
Year	Deep-RL Method	Action	Framework	Code	Ref.
2017	• Learning a policy by fine-tuning a grasp quality CNN via Dex-Net	• Grasping		NA	[78]
2017	• VGG FCNs + Q-learning (vision system takes RGB-D images from multiple views)	• Grasping		https://github.com/andyzeng/apc-vision-toolbox	[89]
2018	• Generative grasping CNN (GG-CNN)	• Grasping		NA	[90]
2018	• DenseNet-FCNs with Q-learning	• Pushing–grasping		https://github.com/andyzeng/visual-pushing-grasping	[119]
2018	• Q-learning with two-stream CNN (ConvNet)	• Picking–placing		https://github.com/andyzeng/arc-robot-vision	[96]
2019	• ResNet with U-net (CNN) + Q-learning	• Suction grasping		NA	[103]
2019	• Place module (FCN ResNet) with Q-learning	• Object-to-placement (suction)		NA	[67]
2019	• Deep Q-learning (DQN)	• Pushing • Grasping		NA	[121]
2019	• -Deep deterministic policy gradients (DDPG) • Actor-critic method	• Pushing • Pick and place		https://k-r-allen.github.io/residual-policy-learning/ https://github.com/k-r-allen/residual-policy-learning	[125]
2019	• DenseNet-FCNs + Q-learning	• Stacking objects		NA	[122]

TABLE 2. *(Continued.)* **Summary of works and their useful links.**

2019	• Training a CNN using Q-learning based on argmax Q(s; a)	• Shifting • Grasping		https://github.com/pantor/learning-shifting-for-grasping	[123]
2020	• DenseNet-FCN + deep Q critic	• Pushing–grasping		https://github.com/choicelab/grasping-invisible	[120]
2020	• Using a single depth image that is fed into a four-branch CNN (has a shared encoder–decoder) • Using circumvent continuous motor control with direct mapping from pixels to Cartesian space inferred from the same depth image	• Grasping		NA	[91]

where they have Synergies between pushing and grasping using DQN algorithm with Mask R-CNN. In terms of learning the nonprehensile rearrangement task in clutter, many feasible frameworks have been presented. For example, Yuan *et al.* [137] learned nonprehensile rearrangement strategy based on Deep Q-network algorithm by exploiting heuristic exploration strategy for reducing the amount of collisions. Whereas, Monte-Carlo Tree Search exploration strategy, which relies on visual inputs coming from an RGB camera, has been presented to learn nonprehensile rearrangement task [118], [138]. Furthermore, Determining where to replace objects inside a cluttered and confined space while rearranging objects to retrieve a target object has been presented in [139]. In contrast to others related work where planning for the placement of removed objects inside a workspace has not received much attention. Rather, removed objects are often placed outside the workspace. In another work, iterative local search (ILS) using heuristics and an e-greedy has been implemented for non-prehensile rearrangement [140], whereby the authors claimed that ILS is equipped with strong heuristics and an e-greedy rollout policy has succeeded at solving various tasks for table-top rearrangement, including a sorting task. In contrast to MCTS, the ILS algorithm is designed to construct a long pushing trajectory that eventually reduces the distance to a goal state.

Teaching agents how to synergize prehensile and nonprehensile actions remains to be a challenge in robotic manipulation, in which various studies have attempted to overcome the limitations of grasping objects in cluttered scenes. Different mechanisms, such as push-to-grasp, shift-to-grasp and suction grasp, have been used. However, performance efficiency still needs to be improved because all of the aforementioned works dealt with specific range and types of objects. In addition, no consideration has been given to the materials of the objects to be grasped (either fragile or deformable). A limitation also exists in terms of the behaviors of synergy types between prehensile and nonprehensile actions that can be executed. Although several researchers trained the nonprehensile policy solely, others attempted to train prehensile and nonprehensile policies by using parallel neural networks. In addition, increasing the complexity of training resources is one of the challenges that needs to be faced during training, and these frameworks require large amounts of data to be able to synergize the two actions. This condition may reversely affect the performance of agents in terms of training time. From the current author's perception, for this particular task, all works on object grasping in cluttered environments attempted to discover and learn synergies between prehensile and nonprehensile actions from experience through model-free deep-RL. On the other hand, model-based learning could produce promising results and can be effectively learned rapidly and efficiently. Another suggestion for improvement is to implement deep networks for Q-function estimation, such as double Q-learning [141], and dueling networks [142] that have the potential to improve performance efficiency.

IV. SIMULATION-TO-REAL-WORLD TRANSFER

Transfer learning from simulation to reality plays a crucial role in robotic applications because this technique requires the consideration of some information about real robots, such as amount of fine tuning, during transfer. Many recent studies have presented promising results in this field. Performing simulation first and then transferring such learning into a real robot is the most effective means to thoroughly understand the training environment, and it might be achieved through a training robot. A learning robot, during simulation, maximizes advantages by collecting sufficient real-world data for robotic learning, but this process can be costly. Furthermore, learning the skills to be adapted to the features of objects being grasped is one of the challenges that robots should overcome. These features might be irrelevant for generalization skills; hence, a robot always selects a small number of relevant features to adapt easily to a specific skill. The robot should be trained with a generalization capability for different object configurations to generate manipulation skills with a wide range of application scenarios. However, generating these scenarios requires numerous training samples, which can be difficult for learning-based algorithms to implement in real robotic systems. Ref. [143] focused on how the learning policy can be deployed in other tasks. By referring to the transfer leaning policy for real robots, Viereck *et al.* [144] trained a neural network to measure object distances in simulation by using a depth sensor (Fig. 10). They aimed to reduce the reality gap on the basis of depth information. The authors in [145], transferred simple learning visuomotor policies of grasping blocks from simulation to a real robot by using progressive networks [146], and Christiano *et al.* [147] trained a fetch robot (Fig. 11) to compensate for several dynamic features, such as friction and physical discrepancies, which might be absent during simulation-based policy learning. In [148], the authors also learned object detection in a complex scenario during simulation training by using visual variations. Then, they transferred the learned policy to a real robot. In all of these aforementioned studies on transfer learning, simulation data were manually generated, and the training policies were replicated for real robot implementation.

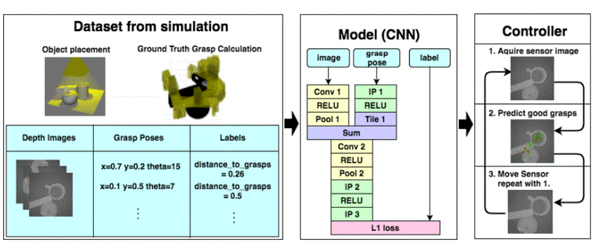


FIGURE 10. Overview of the approach that is divided into three stages, namely, generating training data through simulation, predicting distance-associated nearest grasps via a CNN model and moving the gripper to a predicted grasp position by using a controller [144].

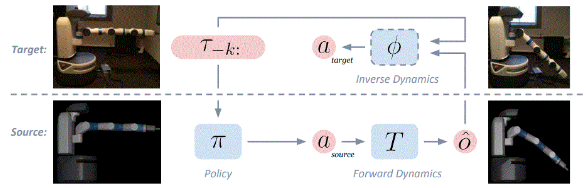


FIGURE 11. Overview of the method applied to a fetch robot in the source simulator (bottom) and target physical world (top) [147].

Another direction has emerged between randomization and supervised learning. James *et al.* [111] aimed to increase the success of transfer learning; they augmented the training during simulation to learn more about grasp predictions of different object shapes. In addition, the Simulated and Physical Articulated Extendable (SPARE) object dataset was generated for use in evaluating different methods in terms of predicting the number and length of links of an articulated object by using a deep neural network (CNN-LSTM), as illustrated in Fig. 12.

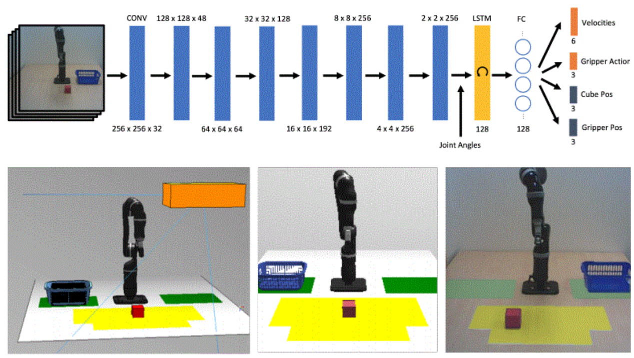


FIGURE 12. Variations in cube, basket and camera positions during training and testing tasks. NN architecture map sequences of the four previous images and joint angles for motor velocities and gripper actions, aside from two extra outputs of the 3D positions of the object and gripper [111].

In another work, task synergy push and grasp policies were trained in simulation by using the modular deep-RL method; the policies were then transferred and applied to real robotic tasks [149]. The asymmetric actor-critic method was presented by using high (partial observation) and low (environment state) dimensions [150], which considerably decreased the number of trainable parameters and increased the critic accuracy during transfer learning. However, the full environment state needs physical parameters to be adaptable to complex dynamic conditions that match a real robot, and obtaining physics parameters is not trivial. Several studies utilized the domain randomization concept for the application of transfer learning, such as transferring policies of deformable objects trained using DDPG with different tasks (e.g. folding a small towel [151]) and transferring in-hand manipulation policies trained using proximal policy optimization (PPO) with the asymmetric actor-critic approach [152]. A learning policy can also be generated via viewpoint-invariant visual servoing using a recurrent CNN (LSTM) by implementing the Monte Carlo method to predict the Q-value of the action [153]. These policies were transferred to real-word robotic applications.

Meanwhile, Chebotar *et al.* [154] studied the optimal distributions of simulation properties by using real-world trajectories to update the simulation parameter distribution during agent training in simulating the task of opening a cabinet drawer and swing-peg-in-hole tasks (Fig. 13). However, the policy was directly learned on randomization rather than domain randomization to learn a randomized-to-canonical adaption function. James *et al.* [155] addressed this problem as their contribution. They implemented Q-function targets via optimization (QT-Opt), which is an off-policy and continuous action generalization of Q-learning, as shown in Fig. 14. In addition, many studies have focused on transfer learning, such as using modular network architectures [156], [157] and randomizing visual appearance and robot dynamics [150], [158].

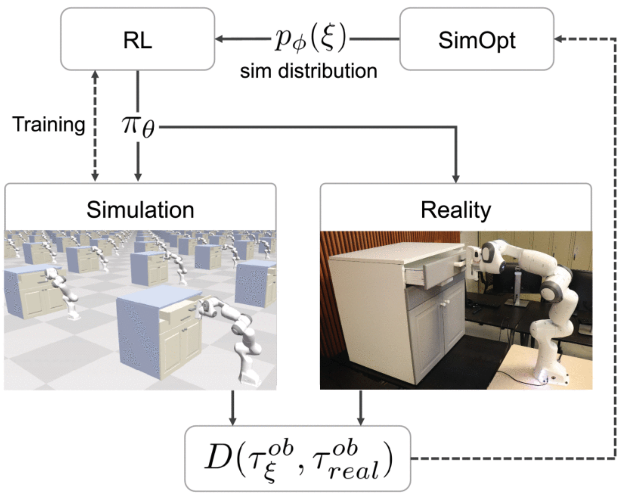


FIGURE 13. Pipeline of simulation optimisation [154].

In accordance with nonprehensile manipulation and rearrangement, such as pushing or shifting, nonprehensile actions have been used in robotics alongside grasping actions because they play an important role in increasing the success rate of grasping when dealing with cluttered and stacked objects. However, using nonprehensile actions requires knowledge on frictional forces; otherwise, robots will have difficulty performing. For example, instead of end-to-end training, the transfer via modularity concept was used in [159] to separate the learned pushing and grasping policies from the raw inputs and outputs. The push policies were trained in

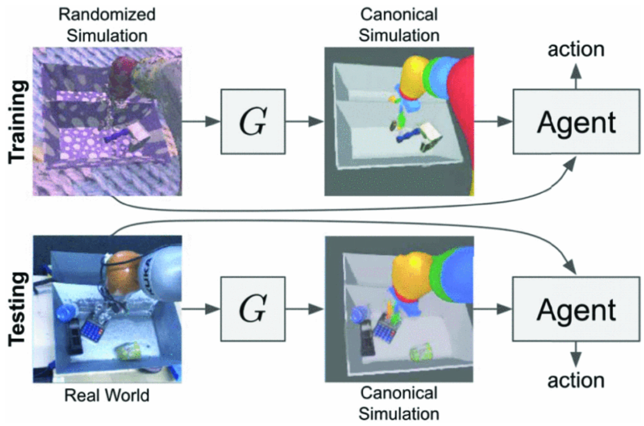


FIGURE 14. Learning to translate randomised simulation images to a chosen canonical simulation version to be fed into the agent [155].

simulation, and the learning was subsequently transferred to a real robot. However, this approach requires augmented reality tags (AR-tags), which are developed in the constrained context of detecting and pushing an object with a robot arm against a uniform green-screen backdrop [160]. A learning push policy to drive mobile robots outdoors was presented in [161]. This policy considers complicated perceptual conditions because it needs different intermediate representations and modules. Moreover, nonprehensile rearrangement involves controlling the robot problem during the interaction with objects for reconfiguring the objects into a predefined target position. Ref. [162] proposed the use of whole arm manipulation to learn how to hold and transport human bodies in rescue and patient care scenarios. The authors used PPO, an actor-critic RL method, to train the policy and directly transfer it to the real-world robotic application. Notably, the system has to forgo access to raw sensor data to avoid the gap between simulation and reality. The authors addressed this issue in their next work [163]. They trained policies end to end by using the deep Q-learning algorithm with CNN, which maps raw pixels as a state-action value then transfers the policy to a real robotic application with supervised examples. Recently, a single universal policy π (a|s, z) was trained by off-policy Q-learning, and the same learned policy was used during testing without any further optimization [164]. However, this approach may not be robust, as explained in [165].

In [166], Arnekvist *et al.* highlighted the problem of transferring knowledge within a family of similar Markov decision processes (MDPs). To solve this problem, they proposed variational policy embedding to learn an adaptable master policy for a family of similar MDPs. Thus, the master policy can adapt to the new family's members rather than finding one robust policy. The policy can also be transferred without a pre-trained dataset. Meanwhile, a cup placing policy was trained using CNN based on Monte Carlo tree search [167]. The purpose was to optimize the augmentation strategy for sim-to-real transfer and enable domain-independent policy learning.

The issue of bridging the reality gap was addressed in [168], which studied how randomized simulated environments and domain adaptation methods could be used to train a grasping system for grasping novel objects from raw monocular RGB images (Fig. 15).

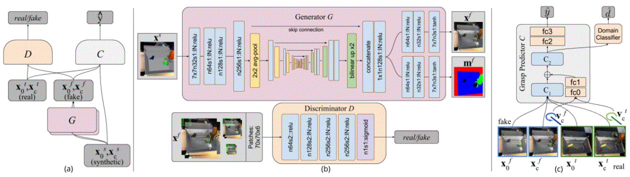


FIGURE 15. Overview of the pixel-level domain adaptation model [168].

In the work by Pedersen *et al.* [169] unknown objects were grasped by training a deep neural network grasping agent on simulated data. Furthermore, domain randomization has been implemented to transfer learned policy from simulation to reality [170], [171]. A method of sim-to-real transfer has been presented in [172], where an end-to-end tactile grasping policy trained in simulation transfers directly to the real world with high fidelity. The work aims to overcome such challenges of the tactile sensing, which could enabled BH-282 Barrett hand through its reinforcement-learned incremental finger closing procedure based on tactile sensory feedback. As methodology, the work used a multi-fingered adaptive tactile (MAT) grasping utilizing deep RL, where MAT models the environment as an MDP defined by $(S, \rho^{\circ}, A, R, T, \gamma, H)$.

V. ROBOTS LEARNING FROM DEMONSTRATION

In the robotics field, learning from demonstration (LfD) is a model where a robot learns new skills by imitating an expert. LfD plays a remarkable role in developing robotics and automation to overcome the limitation of performing complex tasks. In imitating the learning context, the robot acquires the expert's behavioral samples and attempts to execute the task by replicating the expert's actions. In RL, the robot intends to maximize the predicted reward by interacting with the environment. Several studies have concentrated on using prior knowledge from demonstration to initialize a policy [173], [174], and others have attempted to infer the reward function via inverse RL [175], [176]. A few studies focus on improving policies through learning steps [177]–[180]. All of them use human demonstrations to aid in the exploration of learning block-stacking tasks. These methods encounter difficulty in executing assembly tasks that require the collection of multiple demonstrations. Although learning a reward function could partially alleviate this problem, such as the classification of target states stated in [65], an RL-based real-world robotic application has to appoint the target task by means of a manually programmed reward function. It needs to use the same pipeline of perception that the end-to-end method promises to avoid or add extra sensors to check if the task has been successfully executed. Several studies have demonstrated the problem of robots in performing a task in the presence of stationary [181]–[183] or moving obstacles, which are regarded as a special consideration during demonstrations [184]. On the contrary, applying motion planning and task programming is considered a solution for avoiding the complexity of manual programming, and it can be provided by the LfD approach [185].

In addition, high-precision assembly tasks have been the focus of various studies on performing insertion tasks, such as obtaining high-dimensional observations (e.g. joint position, velocity, force, and geometry) [186]–[189]. However, obtaining this information is difficult because it entails complex experiments and supervised learning. Meanwhile, external trajectory planning is one of the techniques to demonstrate assembly tasks [187], [188], but it can be prone to errors in terms of perception. In [190], Schoettler *et al.* presented effectiveness of using LfD with RL compared with using residual RL with the aim of performing an insertion task with minimal prior knowledge, and they did so under noisy conditions. Furthermore, Zhu *et al.* [191] implemented end-to-end RL to learn visuomotor policies without the need for demonstrator actions. However, demonstrations could be involved if raw demonstrator actions are unknown or generated by a different demonstrator. As the states of the environment change, LfD encounters difficulty adapting continuously to the environment's state. Ref. [184] solved this problem by extracting patterns that are important to a given task and can be generalized to different tasks. Obstacle avoidance was considered during demonstrations and used later for model learning. Training a policy to perform book shelving and cloth draping tasks was performed in [65] by using RL with active queries. However, this framework requires additional assumptions for obtaining labels from users, and the authors worked on limited tasks where linear interpolation could be obtained for a state between initial and targeted ones, as mentioned in [192]. Meanwhile, learning of residual tasks has been performed by manually specifying a policy [125], [193]. In [194], collision avoidance in a dynamic environment was presented as a main contribution. The authors used only a few demonstrations, but they were still appropriate for a wide range of obstacles.

Robot learning from demonstration (RLfD) has been progressively implemented on specific tasks with limited environmental conditions, including reduced time and cost, particularly in manufacturing workspaces where the robot needs to avoid stationary obstacles and collision with objects that are shifted and moved by humans by controlling the policy. Therefore, building a control policy from demonstration is required for a robot to avoid collision with movable objects during task execution. Using raw demonstration is also important in hastening the process where learning samples are immediately acquired for the robot so that the robot can perform the action precisely. However, this type of robotic learning is appropriate only for cases with direct contact and a passive controller [195], [196]. Thus, it is not suitable for dealing with multiple DoFs (e.g. robotic arm with multiple DoFs and dual-arm robots). With regard to imitating the learning of robotic manipulation, a generalized capability that allows robots to continuously gain new tasks and exploit pre-existing experience from previous learned tasks is necessary. Piao *et al.* [197] aimed to learn multi-modal imitation demonstrations with different intentions by optimizing a sparse coding lifelong intention dictionary, which enables robots to autonomously imitate various complex behaviors. Their approach was used to develop a lifelong imitation learning framework for inverse reinforcement learning (IRL) to deal with multi-intention imitation learning problems. Specifically, their approach comprises three components: a policy search cost function that incorporates the motion plan trajectory, an efficient initialization of the policy search algorithm with a traditional tracking controller and an NN representation that takes a motion plan as an input and can be trained with RL to track it.

Bed-making task is considered to be difficult home task that can be a challengeable for senior citizens because of reaching motions. Executing bed-making autonomously is a technically challengeable task in terms of perceptive item in an unstructured environment, dealing with sort of deformable objects, sequential decision making, and obstacle avoidance. In [198], they have presented a supervised deep transfer learning approach to locate pick points using depth images for invariance to color and texture, to perform bed-making task. In their approach, they gathered human demonstrations for grasping the sheet and failure detection, by utilizing pre-trained YOLO features in order to facilitate the learning of deep neural network policies. Other works on the execution of folding cloths can be found in [199]–[203]. Instead of improving the synthetic objects to be indistinguishable from real objects, Abolghasemi and Bölöni [204] have trained the vision system to accept synthetic objects as real. They have extended the capabilities of end-to-end LfD architectures to object manipulation in clutter using Variational Autoencoder- Generative Adversarial Network (VAE-GAN). Song *et al.* [205] proposed the use human demonstration and action-view representations to improve learning efficiency. In the work, demonstration is learned from a new low-cost hardware interface that collects grasping demonstrations in clutter, and presenting an end-to-end 6DoF closed loop grasping model with reinforcement learning that transfers to real robots.

VI. WELL-LABELLED DATA

Labelled data comprise a set of samples that have been marked up or annotated with one or more labels. Labelling takes a group of unlabeled data and assigns each sample a meaningful label that is informative. However, data labelling, which has been focused on by many studies, is not a trivial task because it needs an adequate group of samples and an ample amount of time to accomplish such a task with informative labels associated with the target task. Two problems with such a methodology are labelling grasping object locations manually in multiple ways and human labelling, which is biased by semantics. Although several attempts have been made to train agents by trial and error, the samples of labelled data used in such experiments were few, thus making the robot vulnerable to over-fitting errors. Pinto and Gupta [7] collected datasets with a data size of 50 K covering 700 hours, and they generated a large amount of training data that was almost more than one-third the amount of existing data. They trained CNN to predict the grasp orientation by using image patch and self-supervised data in exploiting a cluttered environment rather than a sparse scene. In contrast, Levine *et al.* [112] utilized hand–eye coordination instead of open-loop predictions to continuously observe the gripper and select an appropriate motor command for performing a successful grasp. The advantages of their approach include lack of need for calibration between the agent and camera and training using 800 K grasp attempts on a wider range of objects compared with that in [206] and [7]. In addition, their method requires much time (over two months) to train 14 robots.

Reducing the data collection time is one of the challenges in generating robust robotic grasps and has been the focus of different studies, which initially began with Dex-Net 1.0 [207]. For example, Mahler *et al.* [208] worked on a major extension to generate Dex-Net 2.0 by using synthetic point clouds with robust grasps. They trained a GQ-CNN model by using the cross-entropy method (CEM) of RL to generate point clouds and grasp attempts for predicting robustness. They extended their work to generate a new version of their data called Dex-Net 3.0 [209] that uses synthetic data of depth images for training on bin-picking tasks. However, they adopted synthetic data, which are limited due to the reality gap as mentioned in [210]. Dex-Net 4.0 [211] is a state-of-the-art grasp planner that plans robust grasps for various objects. The method combines the simulation of thousands of 3D object models, analytical wrench mechanics, structured domain randomization and synthetic point clouds to train a deep learning optimization system. The learned policy rapidly processes high-resolution depth images to compute robust robot pick points in various groups of objects for a stationary industrial manipulator. For the implementation of a highly flexible dataset on decluttering surfaces (e.g. homes and machine shops), Staub *et al.* [212] modified Dex-Net 4.0 to generate the Dex-Net MM grasp planner for coping with the parameters of the mobile manipulator because this task can be executed using a mobile manipulator rather than a stationary industrial manipulator; hence, mobile robots were equipped with low-precision sensors and actuators alike. In a surface decluttering experiment where objects were randomly selected from 40 common machine shop objects, the robot was able to recognize, grasp and place the objects into appropriate class bins in 117 out of 135 trials.

One of the difficulties in implementing RL to complex robotic control tasks is the need for a considerable amount of experience in identifying an effective policy for the task at hand. Model based RL is capable of achieving good sample efficiency but needs the ability to learn a dynamics model that is good enough to learn an effective policy. The authors in [213] developed a model-based RL algorithm that combines prior information from preceding tasks with online dynamics model adaptation. Neural networks were used to create and adapt online models that can be used for model-based RL and to learn a control policy that uses iterative linear quadratic regulation. They also used predictive control modelling based on differential dynamics. With respect to data collection, data were collected from different sources to provide a sufficiently diverse dataset. The total training set for the physical robot experiments had data collected at 20 Hz in 6.6 hours [214]. In contrast to Gaussian processes, NNs have constant inference time and tractable training in the Big Data regime and have the potential to represent numerous complex functions, including non-smooth dynamics that are often present in robotics. NNs therefore rely on deterministic models, thereby suffering from overfitting in the early learning stages [215]. Meanwhile, by integrating long-horizon reasoning through RL into a generalizable vision-based framework trained on self-supervised real-world data, [19] attempted to resolve the data limitation caused by the reality gap. Refers. This was not considered by [112] and [7], as they proposed self-supervised grasping systems. Kalashnikov *et al.* [19] suggested, and named QTOpt, an off-policy training approach focused on a continuous-action generalization of Q-learning. They found that a more robust and scalable alternative would be to train only a Q-function and implicitly induce a policy by optimizing this Q-function via stochastic optimization. RL implementations in the real world therefore require significant effort to design and assess the role of reward. While model-free RL approaches have the ability to generalize to new objects [19]and learn tasks such as grasping and pushing through self-supervision [7], [119] pure model-free approaches generally lack the ability to rationalize temporarily extended plans explicitly, making them unsuitable for the problem of learning long-horizon tasks with limited supervision [216].

Data-driven grasp-analysis or grasp planning algorithms for parallel jaw grippers, such as Dex-Net [78], [207], [208], [211], generative grasping convolutional neural network (GG-CNN) [217], Grasp Pose Detection (GPD) [218], viewpoint selection for grasp detection [219] or Fully Convolutional Grasp Quality Convolutional Neural Network (FC-GQ-CNN) [220], typically take sensor input (e.g., an object mesh, a depth-camera image), perform some preprocessing (e.g., image in painting), and produce either a grasp or grasp quality score for a pre-sampled grasp candidate [206]. The majority of these algorithms are based on convolutional neural networks (CNNs) and may be learned from human annotations [80], simulated training data [144], [221], human or self-supervised labels from grasps attempted on a physical system [7], [19], [112], or a combination of the above [168]. Recent work has also explored introducing additional degrees of freedom for grasps in cluttered environments [33], [34], [102], [222], [223], noting that top-down grasps leave out a wide range of feasible high quality grasps on many objects [224]. As for transparent object, there are some works to detect the transparent objects for grasping such Polarized CNN framework, which demonstrates on instance segmentation with Mask R-CNN [225], [226], a grasp quality CNN that takes RGB input [227], and leveraging of deep learning with synthetic training data objects from a single RGB-D image [228].

Selecting appropriate point Grasping of an individual object has been also studied for many years, and there many feasible frameworks have been presented. However, for executing daily life tasks, there is still a challenge to be overcome because the environmental scene in our ordinary life is usually messy, where objects are often significantly influent in term of front and back occlusion, besides stack up and down. In [229], they have mainly focused on grasping plan and selection in a cluttered scene. In another work, segmentation-based framework have been proposed in [230] using Synthetic Data and Mask R-CNN, where object is segmented into primitive shape classes using monocular depth input, with the object to grasp extracted and converted into primitive shape point clouds. In addition, non-Markov picking policies have been presented in [231], which incorporate memory of past failures to modify subsequent actions in robot bin picking. Recently, an end-to-end approach has been proposed in [232], to directly predict the poses, categories and scores (qualities) of all the grasps. They have generated dataset of 23.7k grasps for 79 objects and a multi-object dataset of 20k point clouds with annotations and masks. Furthermore, to overcome the disadvantages of photo-realistic environment simulation, in [233], large-scale dataset has been proposed, which is called Real Embodied Dataset (RED), and it includes full-viewpoint real samples on the upper hemisphere with a modal annotation and enables a simulator that has real visual feedback.

VII. VISION-BASED ROBOTIC GRASP

Executing a complex task from pixel inputs only is considered remarkable and has been focused on by many studies. For example, visual servoing has been widely implemented in many applications associated with robotic tasks [153], [234]. In today's fast-paced world, an increasing number of researchers have examined deep learning to execute complex behaviors from pixels. This area shows promises in performing complex tasks that are difficult to complete [19], [65], [111], [151], [152], [235], [236] especially in goal-conditioned settings [237], [238]. Reach-to-grasp manipulation is performed using DLR, which is made up of many feasible frameworks that help researchers alleviate the challenges in robotic manipulation. Katyal *et al.* [239] proposed a framework using a deep neural network (DNN) that maps a raw image pixel to the Q-value (via the context of Q-learning). Although they made robotic control immune to changes in the robot manipulator or environment during the execution of reaching tasks, their approach needs large amounts of training data, as mentioned in [149] and [155].

In addition, RL, the standard paradigm for solving sequential decision issues, helps robots learn directly from gained experience. However, it is ill-equipped to deal with scalable and uncertain problems during real-world tasks because solving high-level sequential decision tasks depends on the physical robots themselves. Biasing action exploration using state- and action-centric demonstration has been proposed in [240]; it combines object-oriented MDPs (OO-MDPs) and abstract MDPs [241] (AMDPs). However, this approach was trained via trial and error.

Low-level states and actions are challenges in formulating MDPs, and they can be encountered during deep-RL-based robotic manipulation. Such a problem is noticeable in a partially observable MDP (POMDP) where the shape and pose of the object are in a hidden state, such that the images or point clouds produced by sensors cannot store the hidden state. The DQN algorithm has been proposed to overcome this problem. For instance, Gualtieri *et al.* [242] introduced an algorithm that is rather similar to DQN but with a slight difference in the use of a variant of Sarsa rather than Q-learning. They gathered all epochs of experience before labelling the database of experience replay by using the most recent weights of neural networks rather than running a single SGD after each experience. Their target was to overcome the problems of pick–place and re-grasping where the exact geometry of the objects to be handled is unknown (e.g. mugs and bottles). Although they showed a major improvement, they assumed fixed place choices. Meanwhile, Gualtieri and Platt [18] learned 6-DoF grasping and pick–place by using attention focus. The purpose of their work is to generalize these attempts as a single system that can identify 6-DoF grasp and place poses where goal placement is non-trivial.

One of the challenges that can effectively reduce the performance of RL approaches is when the reward can be obtained from a successful trial during the learning of difficult behavioral actions. Learning an object palm-touching task has been performed based on the learning of binocular fixations using either anomaly detection with deep-RL [243] through a weakly supervised, stage-wise learning of simple tasks [244] or without using prior knowledge [245]. In these works, the authors exploited the deep deterministic policy gradient (DDPG) algorithm where the policy and Q-function are approximated by a CNN. Their framework have three stages of learning, namely, fixating an object with cameras, eye–hand coordination by learning to fixate the end effector and using previously acquired skills for an informative shaping reward. However, their work have several limitations. Firstly, the task was executed with minimal consideration of human supervision in terms of kinematic models, calibration parameters or hand-crafted features [246]. Secondly, the detection usually requires processing large amounts of data, a process that is difficult and costly [43]. Avoiding obstacles in 3D space during grasping through multiple DoFs of the robot arm's end effector was presented in [247], where DDPG was implemented to overcome the trajectory planning issue and obstacle avoidance. However, the proposed framework was tested only on the simulation environment. The combination of hindsight experience replay with model-agnostic meta-learning was proposed in [248] to increase the capability of adjusting policies as a key in making learning decisions. Although an improvement in terms of success rate was observed, the framework was only evaluated using the simulation environment.

In unstructured environments, allowing autonomous robots to interact with dynamic objects requires manipulation capabilities that can deal with clutter, changes and variability of objects. To perform robotic tasks a robot must be able to perceive the environmental workspace, prepare and conduct the next action through its sensors. To support us in our everyday tasks, robots need to be able to explore and communicate with unstructured and complex environments found beyond conventional assembly lines and research laboratories. Robust object manipulation is a key component in all robotic applications which require interaction with the environment. In [249], the authors addressed the issue of closed-loop learning policies regarding the combined task of reaching, grasping, and lifting objects. The policies have a mapped depth image in their system, which is collected by a wrist-mounted camera, to the motion of the end-effector, and the gripper opening and closing commands. The authors compared different approaches to RL as a means of improving controller's training efficiency and final performance. Instead of learning to grasp and locate objects using planar manipulation (e.g. a single, demonstrated target state), trust region policy optimization (TRPO) [24] has been introduced as DL. TRPO is close to methods of natural policy gradient and is efficient in optimizing large nonlinear systems, such as neural networks. The perception pipeline used in this work was, however, based on the assumption that objects are positioned on flat surfaces to perform the filtering steps described on the camera images.

The challenge faced by all existing methods involved in the trajectory planning of robotic grasping via supervised learning with a prescribed model prevents the developed grasping strategies from being used for new unknown scenarios. For example, in [250], the authors developed a new method that optimizes grasp trajectories through a policy search method inspired by the success of the RL method. The authors aim to generate a multi-fine robotic hand grasping trajectory and obtain the gripping configuration from objects with a known pose by optimizing the trajectory. Furthermore, research has shown that the softness and flexibility of foam robots provide a great advantage in secure grasping and robust in-hand manipulation [251]. However, working with such a hand requires the application of new modelling and control techniques. Schlagenhauf *et al.* [252] provided users with tools and strategies to create and control dexterous foam robot hands. The primary aim of this work is to evaluate and compare different control strategies for solving the inverse kinematics problem of foam robots. The study also attempted to develop flexible and dexterous foam robot hands that can avoid the IK problem by using a free model-based approach.

However, the authors evaluated their method with simulation as a sufficient tool. In [253], the authors developed a perception and control framework to track a wire cable. The framework relies on a vision-based tactile sensor, GelSight, to estimate the pose of the cable in the grip and the friction forces during cable sliding. Table 3 presents additional related studies on vision-based grasping objects and presents useful information, including deep DL methods, purpose, and limitation.

Learning to grasp unknown objects is not a trivial task to be executed. Thus, selecting suitable grasping algorithms helps researchers to recognize the appropriate algorithm to solve their own problem. The comparative study of different grasping algorithms that suit research results of unknown object grasping has been presented in [271], Furthermore, an overview about recent research issues in robotic grasping and bin picking has also been presented [272], which mainly concentrated on the perception aspects of the problem, associated to computer vision algorithms. Different feasible framework based pixel input has been well studied such as adversarial learning based on ConvNet after AlexNet for effective supervise learning [273], tool use based on task-oriented grasping network (TOG-Net) [274], and grasping deformable objects based on point-level representations [275]. While learning to poke has been executed based on different algorithms such as deep neural networks for modeling the dynamics of robot's interactions directly from images [276], Self-supervised model-based approach using vision-based robotic control [277], and stochastic optimal control with latent representations (SOLAR) [278]. Learning complementary target pre-detection and robotic grasp approach that benefit from each other has been implemented based on Deep Q-Learning Algorithms [279]. However, the framework has only been evaluated in simulation. Zeng *et al.* [280] have proposed end-to-end formulation that jointly learns to infer control parameters for grasping and throwing motion primitives from visual observations (RGB-D images of arbitrary objects in a bin) through trial-and-error. In another work,, framework of leveraging the advantages of active perception has been presented in [281] to perform manipulation tasks. They have used viewpoint changes in determining the object location, which facilitate learning the state representations based on self-supervised concept and performing target directed actions. As they have also compared their framework with vanilla deep Q-learning algorithms as presented in [282]. To learning multimode grasping, Hu *et al.* [283] have presented Policy learning for dynamic grasping algorithm to perform actions such reaching, grasping, and regrasping.

Recently, learning physical object representations for robot manipulation has been performed based on Dense-PhysNet system that actively executes a sequence of dynamic interactions (e.g., sliding and colliding) [284]. Also, Merzic *et al.* [285] have attempted to generate robust grasping under uncertainty based on synthesized control policies that exploit contact sensing, where they have utilized model-free deep reinforcement learning with exploiting Trust Region Policy Optimization (TRPO). Learning-based approaches to grasp planning are preferred over analytical methods due to their ability to better generalize to new, partially observed objects. However, data collection remains one of the biggest bottlenecks for grasp learning methods, particularly for multi-fingered hands. The relatively high dimensional configuration space of the hands coupled with the diversity of objects common in daily life requires a significant number of samples to produce robust and confident grasp success classifiers. In [286], they have presented the first active learning approach to grasping that searches over the grasp configuration space and classifier confidence in a unified manner. In another work, grasp-optimized motion planning (GOMP) has been presented in [287], which is an algorithm that speeds up the execution of a bin-picking robot's operations by incorporating robot dynamics and a set of candidate grasps produced by a grasp planner into an optimizing motion planner. Besides, work in [288], has presented a method that allows end-to-end top-grasp planning methods to generate full six-degree-of-freedom grasps using a single RGBD view as input. Moreover, Yen-Chen *et al.* [75] have proposed a framework called Learning to See before Learning to Act based on directly transferring model parameters from vision networks to affordance prediction networks.

VIII. APPLICATION OF DEEP-RL IN VARIOUS AREAS

This section highlights several applications of deep-RL in different research areas, including assistive robots, pouring of liquids and assembly tasks.

A. HUMANOID ROBOTS

Humanoid-like mobile robots must learn complex motion sequences in human–robot environments so that they can adapt to such motions. The need for assistive robots which are purposely designed and controlled to help the elderly carry out their daily tasks is increasing [289]. Assistive robots are devoted to providing safe grasping in daily tasks. The most crucial challenge that currently faces researchers is how to make the hand of a robot stable and robust during object grasping. In addition, assistive robots can serve as warehouse robots, and the challenge here is still picking up and placing objects in cluttered places or shelves. Achieving tasks in mobile manipulator planning (MMP) often needs thousands of individual motions to be performed (e.g. reasoning about complex targets and feasible movements in configuration space). For example, Chitnis *et al.* [290] exploited the randomized local search algorithm, which can be formulated as an MDP. They aimed to speed up task and motion planning (TAMP), which has the capability to learn samplers of continuous action parameters by exploiting the SGD classifier. Learning samplers for continuous action parameters in TAMP has also been focused on by many other studies [291]–[293]. Although these methods solve the problem of continuous actions in TAMP, they do not consider how learning samplers can be quickly adapted to a new task. The modular

TABLE 3. Other related works, including their methods used, purposes and limitations.

Year	Method	Purpose	Limitations	Ref.
2019	• Genetic algorithm (GA)	• Speeding up the learning agent	• Parameter optimisation	[254]
2019	• Actor-critic using Kronecker-factored trust region (ACKTR)	• Tool use tasks (e.g. picking up a T-shaped tool to reach and drag a target object to a designated region)	• Used only simulation to train and test the model; tried to overcome the issue of using only raw inputs, such as pixel-wise	[50]
2016	• Combining unsupervised learning using deep spatial autoencoders with simple, sample-efficient, trajectory-centric reinforcement learning	• Learning vision-based manipulation skills directly from camera images (e.g. scooping a bag of rice into a bowl with a spatula)	• Demonstrated either on simple synthetic images or without generalisation to new objects	[255]
2019	• Knowledge-induced deep Q-network (KI-DQN)	• Learning to exploit the environment features, such slide-to-wall grasping, where the target object needs to be pushed towards a wall before a feasible grasp	• This approach can be limited to a specific task for particular objects, and fragile flat objects that are grasped using this strategy will be damaged	[256]
2018	• Maximum entropy reinforcement learning algorithm called soft Q-learning (SQL)	• Moving the robot's gripper to a specified target location in Cartesian space (e.g. pushing a cylinder to a specific target and stacking Lego blocks together)	• The placement error still needs to be reduced when the agent is a position-controlled manipulator under the deep reinforcement learning algorithm	[257]
2018	• Guided policy search (GPS) with end-to-end learning	• Performing object placement tasks (e.g. book placement, peg insertion and hanger task)	• This work was implemented only on the simulation environment	[55]
2019	• Actor-duelling-Critic (ADC), which is inspired by the duelling network	• Mitigating the issue of suffering from inaccurate Q estimation resulting in poor performance in a stochastic environment	• Tested on classic gym control environments and obstacle avoidance environments	[258]
2020	• Multi-view approach to closed-loop, end-to-end learning using Q-target optimisation (QT-Opt)	• Mitigating the single-view system issue that sees only one side or view, which results in difficulty in resolving the alignment challenge by combining information from an un-calibrated multiple-camera system and executing good performance on stacking and insertion tasks	• The method was tested using only the Bullet Physics simulator	[259]
2017	• Model-free reinforcement learning algorithms with multilayer neural network representations based on deep deterministic policy gradient (DPPG)	• Presenting a method that allows multiple robots to cooperatively learn • Single policy with deep reinforcement learning (e.g. door opening task)	• This approach entails additional assumptions and limitations as highlighted in [257]	[260]
2019	• Two sample-efficient deep-RL algorithms based on dynamic policy programming: deep P-network (DPN) and duelling deep P-network (DDPN)	• Enhancing sample efficiency and learning • Stability with a few samples by combining smooth policy update with feature extraction in deep neural networks (e.g. robotic cloth manipulation tasks)	• A normal policy lacks these capabilities in principle, including (1) exploration capability to escape the local optima and (2) conservative learning capability not to be trapped in them; these are empirically effective in the field of machine learning	[261]
2019	• Reward–punishment actor-critic (RP-AC) algorithm	• Improving the robot trajectory by acquiring adequate rewards based on the RP-AC framework in robotic non-grasping manipulation tasks (e.g. push task)	• Lack of optimal parameters utilising highly reproducible simulations	[262]
2019	• Path planning via an improved DQN-based learning policy	• Assists the Q network to acquire information in terms of depth experience and supports the model to quickly learn the environmental rules	• Not tested on the robotic grasping task; it was just used for the observation of a variable map environment	[263]
2019	• Guided policy search (GPS) algorithm	• Learning to grasp with primitive-shaped object policies, whereas the policy utilises information from three sources: (1) a vision sensor, (2) the current robot configuration (joint angles and velocities) and (3) the desired grasping posture from a planner	• The framework concentrates on achieving policies for grasping tasks that are not flexible for limited grasping postures	[264]

TABLE 3. *(Continued.)* **Other related works, including their methods used, purposes and limitations.**

2019	• Sequential multitask learning-GPS algorithm (SMT-GPS)	• Operating in sequential multitask learning scenarios due to the lack of GPS algorithms • Ensuring continuous policy learning without catastrophic forgetting. • Tackling the problem of tight coupling between the global and local policies	• Demonstrated the algorithm through learning control policies for two dynamical systems (upward swinging pendulum and peg insertion tasks) and compared it with mirror descent-GPS (MD-GPS) [265]	[266]
2018	• Implementing 'you only look once' (YOLO) to correctly detect the portion of suitable grasp point for stable manipulation. Q-Learning: grasping motion acquisition	• Manipulation of picking up the topmost towels in a folded stack and stacking them on a table	• This approach lacks the smoothness of motion, and more detailed motion planning is required • Suffers when manipulating a deformable linear object (DLO)	[267]
2018	• Sample-efficient, model-based probabilistic inference for learning control (PILCO) [268]	• Resolves the high-dimensional planning problem of DLO manipulation with a reasonable number of samples	• DLO manipulation is constrained in terms of perception, such as non-occlusion in 2D space	[269]
2018	• Tactile sensors and reinforcement learning based on trial and error • Deep neural network classifier: estimating the state of a zipper within a robot's pinch grasp • Contextual multi-armed bandit (C-MAB): maximising cumulative rewards by balancing exploration versus exploitation of the state-action space	• Closure of a zip-lock bag	• Uses discrete actions and constant grasping force	[270]

meta-learning approach was used in [294] to overcome this limitation. Moreover, Chitnis *et al.* trained their model on multiple tasks; their model tried to learn various specializers that can be quickly adapted to a new task by using minimal data.

Silva *et al.* [295] focused on achieving tasks in which task success depends on reaching a target pose that is controlled by a human. They highlighted the issue of human–robot collaboration in scenarios where humans need robot assistance in performing a complex motion to manipulate an object. In this work, the RL-based approach was used to determine a robot's capability to execute a task by using a given current position. However, this current position was indirectly adjusted by prompting the human user. Moreover, they proposed to model indirect control problems via MDP formalism by using trial–error learning based on Q-learning. In [296], the authors proposed two hierarchy plans to perform reach-to-grasp target object with the uncertainty of external perturbations by using the neural-dynamic optimization (ND-Opt) algorithm and low-level RL in operational and joint spaces, respectively. The authors modelled and learned the joint trajectories by using dynamic movement primitives (DMPs); meanwhile, they learned the trajectory with uncertainties via RL by exploiting cascade-forward networks (CF-MNN) and multi-neural network (MNN). However, their approach has a limitation in modelling gravity force because of the tool's weight and direction in coordinate and Cartesian space alike. Su *et al.* [297] utilized two different MNN structures for tool gravity identification on the basis of feed-forward networks (FF-MNN) and cascade-forward networks (CF-MNN) to overcome the limitations in [296]. They built a model that allows improving the performance of nonlinear regression analysis.

Several studies focus on controlling the stability and robustness of hand motion during grasping. For example, computer vision has been used to enhance wrist control in robotized exoskeleton hands to achieve assistive robotic grasping [298]. Although an improvement has been realized in terms of enhancing the controlling part, the challenge of achieving a natural reaching and grasping motion remains. In the field of assistive robots, several studies have attempted to design soft hands. For instance, research has examined the design of a soft hand that has pneumatically enhanced muscles [299] and soft gloves [300] that have a driven cable with the features of flexion and extension that can be custom 3D printed. To decrease stress on the upper limbs, a supernumerary hand with additional force has been designed for grasping [301]. Another assistive robot is the warehouse robot, where a challenge still exists in picking up and placing objects in cluttered places or shelves. Grasping techniques can provide a solution for a particular problem. For example, CNNs based on the eye-in-hand approach is used for object recognition and conventional grippers (e.g. hybrid pinch and suction gripper). Corbato *et al.* [302] discussed several lessons learnt from the Amazon Robotics Challenge. The lessons include the following: task conditions must be the base guide of the solution choice, an individual solution is required for integration and problem solving should be conducted using the hierarchical structure of levels of automation [302]. Evidently, several robotic solutions for the grasping task are associated with human hands [303]. Thus, the real challenge in robots is when the robot gripper needs to plan and navigate in extremely cluttered environments [304]. In contrast to merely storing, kitting needs to prepare products or tools quickly, pick up objects from a cluttered place and place these objects down in cluttered environments by using real-time planning, which can overcome object complexity and detect collision issues [305].

B. HIGH-PRECISION ASSEMBLY TASKS

The authors in [189] aimed at learning policies which could assemble high-precision gear. Through properly interpreting observations they developed robust methods, a method that is more attractive than heuristics or estimating ideal physical dynamics. Their method integrates RL with knowledge about force / torque through the integration of a proper controller for space operation. They also proposed a neural network architecture mirror descent guided policy search, which could be generalised to reasonable environmental variations. However, precise insertion, when using motion planning controllers, is known to fail outside of a very small convergence basin. Previous work also dealt with high-precision assembly tasks, in particular insertion-type tasks. One line of work centered on obtaining high-dimensional observations including geometry, forces, joint positions, and velocities [190]. However, it is difficult to procure this knowledge, thus increasing the difficulty of the experiments and the supervision required. They tried to show how their approach not only solves insertion tasks with far less environmental knowledge but also under noisy conditions. Moreover, the motion planning method and DMP models exhibit stable performance in contact-based tasks but fail if the initial conditions differ [306]. CAD was used in another work [307] for studying robotic assembly. The authors addressed the RL problem, which for learning a control policy relies on random exploration. This condition includes several executions of robots, and is sometimes stuck in solutions that are locally suboptimal. By guiding RL (a policy search algorithm) along a geometric motion path, which is calculated using CAD data, the authors proposed to leverage prior knowledge. Most of these studies, however, considered RGB images obtained in a fixed position with a camera covering the entire scene and several of them trained mappings directly from images to actions. This situation produces very diverse distributions of the image, which leads to difficult learning problems. Thus, learning pose estimation for high-precision robotic assembly using simulated depth images has been proposed in [308].

C. MANIPULATING LIQUIDS

Liquids in human environments are ubiquitous, and appear in many common household activities. Recent robotics research has begun to explore ways robots can think about and manipulate liquids. Several research teams have successfully solved liquid discharge tasks using relatively poor liquid flow physics models [235], [309], [310]. Others have shown that physics-based models have the ability to enhance our understanding of liquid-related actions [311]. A pouring task, for example, involves grasping and moving a container and selecting skills, such as tipping and shaking. The authors suggested stochastic DDP in [312]. When dynamic systems are unknown they used stochastic neural networks to learn. Their proposed method is a type of RL and has three features, namely DDP (graph-structured dynamic systems), model-based RL (hierarchical dynamic systems) and SSA (represents complicated systems). Schenck *et al.* [313] employed granular media rather than liquids for the same purpose. In robotics, however, some studies have used simulators in terms of restricted environments, such as pouring activities, to reason about the liquids. For example, Kunze and Beetz [311] used a simulator to explain the actions of a robot as it attempts to make pancakes, which involves reasoning about the liquid batter. Yamaguchi and Atkeson [312], [314] have used a simulator to rationalize the dissipation of various liquids. These studies, however, used crude liquid simulations for predictive tasks that need no accurate feedback [315]. Yamaguchi and Atkeson followed up their simulated work on pouring using a real robot [316].

IX. CHALLENGES AND FUTURE DIRECTIONS

In term of existing challenges in robotic grasping task, although how robot learning algorithms have been developed and improved to overcome these challenges, there are still many challenges. Fig. 16 shows the current challenges in robotic grasping.

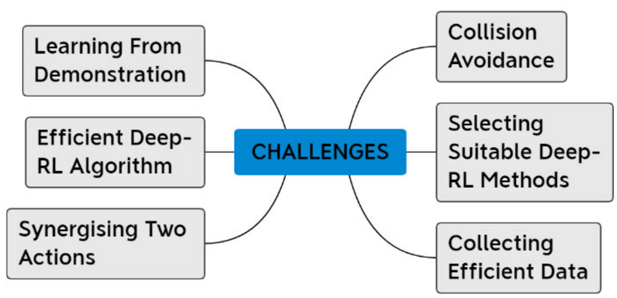


FIGURE 16. **Challenges in robotic grasping.**

A. CHALLENGE OF COLLISION AVOIDANCE

Learning of a good representation using unsupervised learning algorithms (e.g. deep learning) requires large amounts of data, and arbitrary acquired visual representations are usually easily controlled. Although collision avoidance is one of the challenges that robots usually encounter when exclusively performing a robotic task on the intended objects, the environment can be utilized in facilitating robotic tasks, such as grasping an object by pushing the object against the wall or shifting the object to the edge of the table, which serves as an obstacle. Thus, exploiting environmental obstacles sometimes provides another optional strategy when grasping objects from the top (e.g. grasping fragile, flattened, thin objects) is difficult; it remains to be an open challenge in deep-RL. Several studies utilized environmental obstacles [317]–[319], but they exclusively focused on designing a new gripper mechanism by using tactile sensors, which is another gap to be solved using deep-RL.

B. CHALLENGE OF SELECTING SUITABLE DEEP-RL METHODS

Another challenge in deep-RL is determining the most suitable deep-RL methods for manipulation purposes. A number of model-free and off-policy deep-RL methods, such as Q-learning, Monte Carlo, corrected Monte Carlo (Corr-MC), DDPG, PCL and DQL, have been proposed to solve off-policy tasks in the conceptual context of Q-learning. Quillen *et al.* [282] studied different RL algorithms to determine which off-policy RL algorithms are the most suitable for vision-based robotic grasping. Their evaluation indicated that DQL performs better in grasping tasks than other algorithms do in low-data regimes for off-policy and on-policy learning. Meanwhile, DQL, PCL and Corr-MC methods are stable, whereas DDPG is unstable. Although model-free deep-RL has produced promising results in domains ranging from video games to simulated robotic manipulation and locomotion, model-free methods are known to perform poorly because the interaction time with the environment is limited, as is the case for most real-world robotic tasks. With regard to the application of RL in the control field, the first consideration should be the action space because the majority of previous RL methods are applicable in domains with discrete actions, which are associated with value function estimation. Although RL can be useful in performing robotic tasks, which are difficult to model, not all RL algorithms are suitable for all types of robotic manipulation. For example, the tabular RL algorithm has limitations in terms of capacity; it requires a model for the training of each map, and the model has no generalization performance. In the current form of guided policy search (GPS) [264], the GPS algorithm cannot be implemented in sequential multitask learning scenarios because of its batch-style training requirement, where all training samples are collectively provided at the start of the learning process.

C. CHALLENGE OF COLLECTING EFFICIENT DATA

Collecting efficient data is currently the key in performing complex manipulating tasks and increases the opportunity of iteration success during task execution. Although enhancing real data with synthetic data improves the success rate, this augmentation can increase the amount of data storage, which adversely affects the performance of running data in terms of required time. These approaches can be utilized with simulators as well. To collect more data, researchers have trained more than one robot to help them collect data; this method is effective in gathering sufficient data, as demonstrated in [7] and [206]. In this case, collecting data requires considerable time and large memory or storage. Thus, the gap between real and synthetic data should be reduced, and simulated data play a crucial role in reducing this gap. Several studies attempted to address this gap by reducing the difference through the learning process during the execution of grasping tasks. Another attempted to create a map from synthetic to real data by using deep learning. In addition, progressive networks have been proposed to bridge such a gap by using transfer learning from low-level to high-level visual features for new tasks. However, all data, whether real or synthetic, that have been used for specific tasks cannot be useful in certain domains because data are particularly generated for a certain type of robots with specific configurations. Thus, we still need a method that can transform data and be widely used in various platform robots and configurations. Moreover, producing an efficient algorithm remains to be an open challenge. To the best of the current author's knowledge, the model-based method might be able to overcome the limits in data efficiency.

D. CHALLENGE OF LEARNING FROM DEMONSTRATION

With regard to the challenges in LfD, an alternative strategy for dealing with the data demand is to train in simulation and transfer the learned controller to real hardware or to augment real-world training with synthetic data. To transfer RL in robotics, most RL studies employ the following research paths: pre-training an RL model in simulators, transferring the model to robots and fine tuning the model parameters. These processes are usually executed sequentially; that is, after the RL models have been pre-trained and transferred to robots, no meaningful experience or knowledge from the simulators can be provided to the final models fine-tuned on real-life robots. Thus, training robots directly in the real environment is unsafe, and training in simulation and deploying in the real world have become a common trend in robotics under the theme of sim-to-real transfer. An important first step to sim-to-real transfer is sim-to-sim transfer [165]. Numerous recent studies have examined the transfer of policies across different simulation environments, across dynamic models and from simulation to real environments [320]. Furthermore, the exploration method in robotics is still an open challenge, particularly in finding an effective method to deal with a continuous high-dimensional action space. Although the ϵ greedy strategy exhibits a significant improvement, it suffers from several issues, including dealing equivalently with random execution of actions and lack of exploration workspace. In addition, the off-policy is limited to a specific task that cannot be used for more than one platform because it can be generated by training the agent in performing a particular task. As for the on-policy, it can execute robotic tasks more efficiently than the off-policy can in terms of updating the policy continuously; however, it still depends on initial conditions and training steps, thus focusing more on exploitation than on exploration. Hence, the on-policy is restricted to local optimization. The deterministic policy adds noise to actions during training time, such that reducing the scale of noise might assist in obtaining high-quality training time. However, this policy is still inadequate for facing the problem of sparse and deceptive rewards. Another challenge in training a policy through the exploration strategy is that its performance differs in terms of the environment's and robot's configurations, which causes the exploration strategy to face with difficulty in measuring the extent of successful improvement. This challenge will place real robots in unsafe conditions, particularly when exploring with uncertainty by using fragile robots.

E. CHALLENGE OF EFFICIENT DEEP-RL ALGORITHM

Another challenge remains in real robotic applications, and that is the need for an efficient deep-RL algorithm to overcome the limitations of real-world obstacles, including learning fast and efficiently. Certain variables in learning robotic tasks need to be improved in different fields, namely, (1) model-based learning, (2) learning from prior experience (replay experience), (3) transfer learning and (4) domain adaptation. Such an improvement is a new direction for researchers in increasing the efficiency of executing robotic tasks. For example, the deep network architecture can be used to successfully predict over 100 steps of future frames because this approach can be trained visually and potentially applied to other visually rich RL problems. Several approaches that are beneficial in predicting several hundred frames despite being trained to estimate 10 future frames have also been introduced; these include stochastic adversarial video prediction [321], generative adversarial network [322] and variational autoencoder variants [323]. In deep-RL, the agent still learns from the same task procedures, and learning from other tasks remains difficult. To the best of the current author's knowledge, a gap exists between deep-RL algorithms and humans in terms of learning to perform tasks. Human knowledge is a cumulative process, and deep-RL algorithms derive inspiration from humans in terms of learning based on either trial and error or replay experience. However, the gap in learning style between humans and robots remains to be a challenge in terms of learning from other tasks. For example, humans can gather all prior knowledge to perform a novel task with fewer trials compared with robots that need to gather considerable data during training time until they can adapt to a new task. Robots learn and train on a specific task with limitations in the environment's and robot's configurations, and they cannot use prior knowledge on other tasks when performing novel ones. The most recent work on alleviating this gap is model-based learning, which is a promising approach in achieving fast and efficient learning.

F. CHALLENGE OF SYNERGISING TWO ACTIONS

Teaching agents how to synergize between prehensile and nonprehensile actions remains to be a challenge in robotic manipulation, although various studies have been conducted to overcome the limitations of grasping objects in cluttered scenes. Different mechanisms, such as push to grasp, shift to grasp and suction grasp, have been utilized. However, the performance efficiency of these mechanisms still needs to be improved. The aforementioned studies focus on dealing with specific range types of objects, and no consideration has been given to the materials of the objects to be grasped (fragile or deformable). In addition, a limitation exists in terms of the behavior of synergy types between prehensile and nonprehensile actions that can be executed. Several studies trained only a nonprehensile policy, whereas others attempted to train prehensile and nonprehensile policies by using parallel neural networks. Increasing the complexity of training resources is one of the challenges encountered during training, and these frameworks require a large amount of data to be able to synergize two actions. This requirement may adversely affect the performance of agents in terms of training time. In this particular task, all studies on object grasping in cluttered environments have attempted to discover and learn synergies between prehensile and nonprehensile actions from experience through model-free deep-RL. Meanwhile, model-based learning can produce promising results and can learn rapidly and efficiently. Another suggestion for improvement is to implement deep networks in Q-function estimation. Examples of these networks are double Q-learning [141] and dueling networks [142], which have the potential to improve performance efficiency.

X. CONCLUSION

This article discusses different topics, including grasping in clutter, transfer of the learned policy from simulation to a real robot, learning from demonstration, learning from pixel input and learning other tasks (pouring liquid and humanoid robot-based grasping), which are all associated with robotic grasping. The review also highlights different strategies of data collection and model types that show how speedily robots can learn transition models and how these models can be generalized to new tasks. In today's fast-paced world, deep-RL-based robotic manipulation is becoming increasingly crucial in alleviating the problems in performing complex robotic tasks. In the last decade, many researchers attempted to achieve complex robotic performance by teaching agents to execute various tasks autonomously through different approaches, including robotic learning based on trial and error, end-to-end learning and vision-based robotic grasping. This article also presents the challenges in robotic grasping and how robot learning algorithms have been developed and improved to overcome these challenges.

Introducing the deep-RL approach to robotics has considerably improved the learning features of the environment and agent and the main elements of the RL concept. Given the increasing demands for the realization of complex tasks, deep learning has been combined with RL to transform robotic learning by utilizing most of the features that can be useful for robotic tasks in different robotic applications. The main contribution of this review is that it focuses on deep-RL-based object grasping, including applicable approaches and their applications, limitations and existing challenges. It also provides recommendations that are based on the summary of all relevant studies in this field. This review can be valuable and applied to recent work on robotic grasping because it presents the challenges and future research directions in the robotic grasping field. To the best of the current author's knowledge, robotic grasping is still a challenging topic despite the extensive work performed on it. Different research directions can be pursued in the future. These directions include (1) efficient RL, (2) long-horizon reasoning, (3) hierarchical RL, (4) meta-RL, (5) reward function, (6) multi-model, (7) lifelong learning and (8) simulation-to-reality concept.

REFERENCES

[1] G. Garcia, J. Corrales, J. Pomares, and F. Torres, "Survey of visual and force/tactile control of robots for physical interaction in Spain," *Sensors*, vol. 9, no. 12, pp. 9689–9733, Dec. 2009, doi: 10.3390/s91209689.

[2] C. Chi, X. Sun, N. Xue, T. Li, and C. Liu, "Recent progress in technologies for tactile sensors," *Sensors*, vol. 18, no. 4, p. 948, Apr. 2018, doi: 10.3390/s18040948.

[3] L. Zou, C. Ge, Z. J. Wang, E. Cretu, and X. Li, "Novel tactile sensor technology and smart tactile sensing systems: A review," *Sensors*, vol. 17, no. 11, pp. 1–24, Nov. 2017, doi: 10.3390/s17112653.

[4] S. Luo, J. Bimbo, R. Dahiya, and H. Liu, "Robotic tactile perception of object properties: A review," *Mechatronics*, vol. 48, pp. 54–67, Nov. 2017, doi: 10.1016/j.mechatronics.2017.11.002.

[5] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255–260, Jul. 2015, doi: 10.1126/science.aaa8415.

[6] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[7] L. Pinto and A. Gupta, "Supersizing self-supervision: Learning to grasp from 50K tries and 700 robot hours," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Stockholm, Sweden, Jun. 2016, pp. 3406–3413, doi: 10.1109/ICRA.2016.7487517.

[8] P.-C. Yang, K. Sasaki, K. Suzuki, K. Kase, S. Sugano, and T. Ogata, "Repeatable folding task by humanoid robot worker using deep learning," *IEEE Robot. Autom. Lett.*, vol. 2, no. 2, pp. 397–403, Apr. 2017, doi: 10.1109/LRA.2016.2633383.

[9] D. Hossain, G. Capi, M. Jindai, and S.-I. Kaneko, "Pick-place of dynamic objects by robot manipulator based on deep learning and easy user interface teaching systems," *Ind. Robot, Int. J.*, vol. 44, no. 1, pp. 11–20, Jan. 2017, doi: 10.1108/IR-05-2016-0140.

[10] S. Kumra and C. Kanan, "Robotic grasp detection using deep convolutional neural networks," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Vancouver, BC, Canada, Dec. 2017, pp. 769–776, doi: 10.1109/IROS.2017.8202237.

[11] G. de Magistris, A. Munawar, and P. Vinayavekhin, "Teaching a robot pick and place task using recurrent neural network," in *Proc. ViEW*, Yokohama, Japan, Dec. 2016, pp. 1–4.

[12] D. Hossain, G. Capi, and M. Jindai, "Optimizing deep learning parameters using genetic algorithm for object recognition and robot grasping," *J. Electron. Sci. Technol.*, vol. 16, no. 1, pp. 11–15, 2018, doi: 10.11989/JEST.1674-862X.61103113.

[13] D. Hossain and G. Capi, "Multiobjective evolution for deep learning and its robotic applications," in *Proc. 8th Int. Conf. Inf., Intell., Syst. Appl. (IISA)*, Larnaca, Cyprus, Mar. 2017, pp. 1–6, doi: 10.1109/IISA.2017.8316404.

[14] D. Hossain and G. Capi, "Application of deep belief neural network for robot object recognition and grasping," in *Proc. 2nd IEEJ Int. Workshop Sens., Actuation, Motion Control (SAMCON)*, Tokyo, Japan, Mar. 2016, pp. 1–4.

[15] S. Bhagat and H. Banerjee, "Deep reinforcement learning for soft, flexible robots: Brief reviewwith impending challenges," *Robotics*, vol. 8, no. 4, pp. 1–36, 2019, doi: 10.3390/robotics8010004.

[16] M. Ersen, E. Oztop, and S. Sariel, "Cognition-enabled robot manipulation in human environments: Requirements, recent work, and open problems," *IEEE Robot. Autom. Mag.*, vol. 24, no. 3, pp. 108–122, Sep. 2017, doi: 10.1109/MRA.2016.2616538.

[17] V. François-lavet, P. Henderson, R. Islam, M. G. Bellemare, and J. Pineau, "An introduction to deep reinforcement learning," *Found. Trends Mach. Learn.*, vol. 11, nos. 3–4, pp. 219–354, 2018, doi: 10.1561/2200000071.

[18] M. Gualtieri and R. Platt, "Learning 6-DoF grasping and pick-place using attention focus," 2018, pp. 1–10, *arXiv:1806.06134*. [Online]. Available: http://arxiv.org/abs/1806.06134

[19] D. Kalashnikov, A. Irpan, P. Pastor, J. Ibarz, A. Herzog, E. Jang, D. Quillen, E. Holly, M. Kalakrishnan, V. Vanhoucke, and S. Levine, "QT-OPT: Scalable deep reinforcement learning for vision-based robotic manipulation," 2018, pp. 1–23, *arXiv:1806.10293*. [Online]. Available: http://arxiv.org/abs/1806.10293

[20] T. Zhao, M. Deng, Z. Li, and Y. Hu, "Cooperative manipulation for a mobile dual-arm robot using sequences of dynamic movement primitives," *IEEE Trans. Cognit. Develop. Syst.*, vol. 12, no. 1, pp. 18–29, Mar. 2020, doi: 10.1109/TCDS.2018.2868921.

[21] S. Krishnan, A. Garg, R. Liaw, B. Thananjeyan, L. Miller, F. T. Pokorny, and K. Goldberg, "SWIRL: A sequential windowed inverse reinforcement learning algorithm for robot tasks with delayed rewards," *Int. J. Robot. Res.*, vol. 38, nos. 2–3, pp. 126–145, Jul. 2019, doi: 10.1177/0278364918784350.

[22] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," 2015, pp. 1–14, *arXiv:1509.02971*. [Online]. Available: http://arxiv.org/abs/1509.02971

[23] N. Heess, D. TB, S. Sriram, J. Lemmon, J. Merel, G. Wayne, Y. Tassa, T. Erez, Z. Wang, S. M. Ali Eslami, M. Riedmiller, and D. Silver, "Emergence of locomotion behaviours in rich environments," 2017, pp. 1–14, *arXiv:1707.02286*. [Online]. Available: http://arxiv.org/abs/1707.02286

[24] J. Schulman, J. Eecs, B. Edu, P. Abbeel, P. Cs, and B. Edu, "Trust Region Policy Optimization," in *Proc. 31st Int. Conf. Mach. Learn.*, Lille, France, Jul. 2015, pp. 1889–1897.

[25] V. Mnih, M. Mirza, A. Graves, T. Harley, T. P. Lillicrap, and D. Silver, "Asynchronous methods for deep reinforcement learning," in *Proc. 33rd Int. Conf. Mach. Learn.*, New York, NY, USA, Jun. 2016, pp. 1928–1937.

[26] R. V. Martinez, J. L. Branch, C. R. Fish, L. Jin, R. F. Shepherd, R. M. D. Nunes, Z. Suo, and G. M. Whitesides, "Robotic tentacles with three-dimensional mobility based on flexible elastomers," *Adv. Mater.*, vol. 25, no. 2, pp. 205–212, Jan. 2013, doi: 10.1002/adma.201203002.

[27] A. A. Stokes, R. F. Shepherd, S. A. Morin, F. Ilievski, and G. M. Whitesides, "A hybrid combining hard and soft robots," *Soft Robot.*, vol. 1, no. 1, pp. 70–74, Jul. 2014, doi: 10.1089/soro.2013.0002.

[28] D. Guo, F. Sun, T. Kong, and H. Liu, "Deep vision networks for real-time robotic grasp detection," *Int. J. Adv. Robot. Syst.*, vol. 14, no. 1, pp. 1–8, Dec. 2016, doi: 10.1177/1729881416682706.

[29] J. Zhou, S. Chen, and Z. Wang, "A soft-robotic gripper with enhanced object adaptation and grasping reliability," *IEEE Robot. Autom. Lett.*, vol. 2, no. 4, pp. 2287–2293, Oct. 2017, doi: 10.1109/LRA.2017.2716445.

[30] H. A. Pierson and M. S. Gashler, "Deep learning in robotics: A review of recent research," *Adv. Robot.*, vol. 31, no. 16, pp. 821–835, Aug. 2017, doi: 10.1080/01691864.2017.1365009.

[31] Y. Li, Q. Lei, C. Cheng, G. Zhang, W. Wang, and Z. Xu, "A review: Machine learning on robotic grasping," in *Proc. SPIE, 11th Int. Conf. Mach. Vis. (ICMV)*, Munich, Germany, vol. 11041, Mar. 2019, Art. no. 110412U, doi: 10.1117/12.2522945.

[32] S. Bhagat, H. Banerjee, Z. Ho Tse, and H. Ren, "Deep reinforcement learning for soft, flexible robots: Brief review with impending challenges," *Robotics*, vol. 8, no. 1, p. 4, Jan. 2019, doi: 10.3390/robotics8010004.

[33] Z. Kappassov, J.-A. Corrales, and V. Perdereau, "Tactile sensing in dexterous robot hands—Review," *Robot. Auto. Syst.*, vol. 74, pp. 195–220, Dec. 2015, doi: 10.1016/j.robot.2015.07.015.

[34] A. Saudabayev and H. A. Varol, "Sensors for robotic hands: A survey of state of the art," *IEEE Access*, vol. 3, pp. 1765–1782, Sep. 2015, doi: 10.1109/ACCESS.2015.2482543.

[35] Y. Li, J. Krahn, and C. Menon, "Bioinspired dry adhesive materials and their application in robotics: A review," *J. Bionic Eng.*, vol. 13, no. 2, pp. 181–199, 2016, doi: 10.1016/S1672-6529(16)60293-7.

[36] B. Gorissen, D. Reynaerts, S. Konishi, K. Yoshida, J. W. Kim, and M. De Volder, "Elastic Inflatable Actuators for Soft Robotic Applications," *Adv. Mater.*, vol. 29, no. 43, pp. 1–14, 2017, doi: 10.1002/adma.201604977.

[37] P. Polygerinos, N. Correll, S. A. Morin, B. Mosadegh, C. D. Onal, K. Petersen, M. Cianchetti, M. T. Tolley, and R. F. Shepherd, "Soft robotics: Review of fluid-driven intrinsically soft devices; manufacturing, sensing, control, and applications in human-robot interaction," *Adv. Eng. Mater.*, vol. 19, no. 12, pp. 1–22, 2017, doi: 10.1002/adem.201700016.

[38] J. Shintake, V. Cacucciolo, D. Floreano, and H. Shea, ''Soft robotic grippers,'' *Adv. Mater.*, vol. 30, no. 29, Jul. 2018, Art. no. 1707035, doi: 10.1002/adma.201707035.
[39] A. Sahbani, S. El-Khoury, and P. Bidaud, ''An overview of 3D object grasp synthesis algorithms,'' *Robot. Auto. Syst.*, vol. 60, no. 3, pp. 326–336, Mar. 2012, doi: 10.1016/j.robot.2011.07.016.
[40] J. Kober, J. A. Bagnell, and J. Peters, ''Reinforcement learning in robotics: A survey,'' *Int. J. Robot. Res.*, vol. 32, no. 11, pp. 1238–1274, Sep. 2013, doi: 10.1177/0278364913495721.
[41] J. Bohg, A. Morales, T. Asfour, and D. Kragic, ''Data-driven grasp synthesis—A survey,'' *IEEE Trans. Robot.*, vol. 30, no. 2, pp. 289–309, Apr. 2014, doi: 10.1109/TRO.2013.2289018.
[42] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, ''Deep reinforcement learning: A brief survey,'' *IEEE Signal Process. Mag.*, vol. 34, no. 6, pp. 26–38, Nov. 2017.
[43] N. Khanzhahi, B. Masoumi, and B. Karasfi, ''Deep reinforcement learning issues and approaches for the multi-agent centric problems,'' in *Proc. the 9th Conf. Artif. Intell. Robot. 2nd Asia–Pacific Int. Sym. (AIAR)*, Kish Island, Iran, Aug. 2019, pp. 87–95, doi: 10.1109/AIAR.2018.8769770.
[44] V. François-Lavet, P. Henderson, R. Islam, M. G. Bellemare, and J. Pineau, ''An introduction to deep reinforcement learning,'' *Found. Trends Mach. Learn.*, vol. 11, nos. 3–4, pp. 219–354, 2018, doi: 10.1561/2200000071.
[45] Y. Li, ''Deep reinforcement learning: An overview,'' 2017, pp. 1–85, *arXiv:1701.07274*. [Online]. Available: http://arxiv.org/abs/1701.07274
[46] Q. Jia, J. Cai, Z. Cao, Y. Wu, X. Zhao, and J. Yu, ''Deep learning for object detection and grasping: A survey,'' in *Proc. IEEE Int. Conf. Inf. Autom. (ICIA)*, Wuyishan, China, Aug. 2018, pp. 427–432, doi: 10.1109/ICInfA.2018.8812318.
[47] S. Caldera, A. Rassau, and D. Chai, ''Review of deep learning methods in robotic grasp detection,'' *Multimodal Technol. Interact.*, vol. 2, no. 3, pp. 57–81, 2018, doi: 10.3390/mti2030057.
[48] N. Sünderhauf, O. Brock, W. Scheirer, R. Hadsell, D. Fox, J. Leitner, B. Upcroft, P. Abbeel, W. Burgard, M. Milford, and P. Corke, ''The limits and potentials of deep learning for robotics,'' *Int. J. Robot. Res.*, vol. 37, nos. 4–5, pp. 405–420, Apr. 2018, doi: 10.1177/0278364918770733.
[49] K. Chatzilygeroudis, V. Vassiliades, F. Stulp, S. Calinon, and J.-B. Mouret, ''A survey on policy search algorithms for learning robot controllers in a handful of trials,'' *IEEE Trans. Robot.*, vol. 36, no. 2, pp. 328–347, Apr. 2020, doi: 10.1109/TRO.2019.2958211.
[50] H. Nguyen and H. La, ''Review of deep reinforcement learning for robot manipulation,'' in *Proc. 3rd IEEE Int. Conf. Robot. Comput. (IRC)*, Naples, Italy, Feb. 2019, pp. 590–595, doi: 10.1109/IRC.2019.00120.
[51] G. Du, K. Wang, S. Lian, and K. Zhao, ''Vision-based robotic grasp detection from object localization, object pose estimation to grasp estimation: A review,'' 2019, pp. 1–35, *arXiv:1905.06658*. [Online]. Available: https://arxiv.org/abs/1905.06658
[52] O. Kroemer, S. Niekum, and G. Konidaris, ''A review of robot learning for manipulation: Challenges, representations, and algorithms,'' 2019, pp. 1–43, *arXiv:1907.03146*. [Online]. Available: http://arxiv.org/abs/1907.03146
[53] F. den Hengst, E. M. Grua, A. el Hassouni, and M. Hoogendoorn, ''Reinforcement learning for personalization: A systematic literature review,'' in *Proc. Data Sci.*, Apr. 2020, pp. 1–41, doi: 10.3233/DS-200028.
[54] P. Ardón, È. Pairet, K. S. Lohan, S. Ramamoorthy, and R. P. A. Petrick, ''Affordances in robotic tasks—A survey,'' 2020, pp. 1–18, *arXiv:2004.07400*. [Online]. Available: http://arxiv.org/abs/2004.07400
[55] W. Guo, C. Wang, Y. Fu, and F. Zha, ''Deep reinforcement learning algorithm for object placement tasks with manipulator,'' in *Proc. IEEE Int. Conf. Intell. Saf. Robot. (ISR)*, Shenyang, China, Sep. 2018, pp. 608–613.
[56] X. Wang, X. Jiang, J. Zhao, S. Wang, and Y.-H. Liu, ''Grasping objects mixed with towels,'' *IEEE Access*, vol. 8, pp. 129338–129346, Jul. 2020, doi: 10.1109/ACCESS.2020.3008763.
[57] M. Yan, Y. Zhu, N. Jin, and J. Bohg, ''Self-supervised learning of state estimation for manipulating deformable linear objects,'' *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 2372–2379, Apr. 2020, doi: 10.1109/LRA.2020.2969931.
[58] Y. Deng, X. Guo, Y. Wei, K. Lu, B. Fang, D. Guo, H. Liu, and F. Sun, ''Deep reinforcement learning for robotic pushing and picking in cluttered environment,'' in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Macau, China, Jan. 2019, pp. 619–626, doi: 10.1109/IROS40897.2019.8967899.
[59] B. Wu, I. Akinola, and P. K. Allen, ''Pixel-attentive policy gradient for multi-fingered grasping in cluttered scenes,'' in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Macau, China, Jan. 2019, pp. 1789–1796, doi: 10.1109/IROS40897.2019.8968263.
[60] D. Morrison, P. Corke, and J. Leitner, ''Multi-view picking: Next-best-view reaching for improved grasping in clutter,'' in *Proc. Int. Conf. Robot. Autom. (ICRA)*, Montreal, QC, Canada, Aug. 2019, pp. 8762–8768, doi: 10.1109/ICRA.2019.8793805.
[61] Y. Xiao, S. Katt, A. Ten Pas, S. Chen, and C. Amato, ''Online planning for target object search in clutter under partial observability,'' in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Montreal, QC, Canada, Aug. 2019, pp. 8241–8247, doi: 10.1109/ICRA.2019.8793494.
[62] J. K. Li, D. Hsu, and W. S. Lee, ''Act to see and see to act: POMDP planning for objects search in clutter,'' in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Dec. 2016, pp. 5701–5707, doi: 10.1109/IROS.2016.7759839.
[63] E. Jang, C. Devin, V. Vanhoucke, and S. Levine, ''Grasp2Vec: Learning object representations from self-supervised grasping,'' 2018, pp. 1–15, *arXiv:1811.06964*. [Online]. Available: http://arxiv.org/abs/1811.06964
[64] M. Graña, M. Alonso, and A. Izaguirre, ''A panoramic survey on grasping research trends and topics,'' *Cybern. Syst.*, vol. 50, no. 1, pp. 40–57, Jan. 2019, doi: 10.1080/01969722.2018.1558013.
[65] A. Singh, L. Yang, K. Hartikainen, C. Finn, and S. Levine, ''End-to-end robotic reinforcement learning without reward engineering,'' 2019, pp. 1–14, *arXiv:1904.07854*. [Online]. Available: http://arxiv.org/abs/1904.07854
[66] S. Hasegawa, K. Wada, K. Okada, and M. Inaba, ''Detecting and picking of folded objects with a multiple sensor integrated robot hand,'' in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Madrid, Spain, Oct. 2018, pp. 1138–1145, doi: 10.1109/IROS.2018.8593398.
[67] K. Zakka, A. Zeng, J. Lee, and S. Song, ''Form2Fit: Learning shape priors for generalizable assembly from disassembly,'' 2019, pp. 1–7, *arXiv:1910.13675*. [Online]. Available: http://arxiv.org/abs/1910.13675
[68] D. Prattichizzo and J. C. Trinkle, ''Grasping,'' in *Springer Handbook Robotics. Springer Handbooks*, B. Siciliano and O. Khatib, Eds. Cham, Switzerland: Springer, 2008, pp. 671–700, doi: 10.1007/978-3-319-32552-1_38.
[69] J. Weisz and P. K. Allen, ''Pose error robust grasping from contact wrench space metrics,'' in *Proc. Int. Conf. Robot. Autom. (ICRA)*, Saint Paul, MN, USA, Jun. 2012, pp. 557–562, doi: 10.1109/ICRA.2012.6224697.
[70] F. T. Pokorny and D. Kragic, ''Classical grasp quality evaluation: New algorithms and theory,'' in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Tokyo, Japan, Jan. 2014, pp. 3493–3500, doi: 10.1109/IROS.2013.6696854.
[71] A. Rodriguez, M. T. Mason, and S. Ferry, ''From caging to grasping,'' *Int. J. Robot. Res.*, vol. 31, no. 7, pp. 886–900, Jun. 2012, doi: 10.1177/0278364912442972.
[72] R. Balasubramanian, L. Xu, P. D. Brook, J. R. Smith, and Y. Matsuoka, ''Physical human interactive guidance: Identifying grasping principles from human-planned grasps,'' *IEEE Trans. Robot.*, vol. 28, no. 4, pp. 899–910, Aug. 2012, doi: 10.1109/TRO.2012.2189498.
[73] L. Pinto and A. Gupta, ''Learning to push by grasping: Using multiple tasks for effective learning,'' in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Singapore, May 2017, pp. 2161–2168, doi: 10.1109/ICRA.2017.7989249.
[74] F. Ebert, S. Dasari, A. X. Lee, S. Levine, and C. Finn, ''Robustness via retrying: Closed-loop robotic manipulation with self-supervised learning,'' 2018, pp. 1–12, *arXiv:1810.03043*. [Online]. Available: http://arxiv.org/abs/1810.03043
[75] L. Yen-Chen, A. Zeng, S. Song, P. Isola, and T.-Y. Lin, ''Learning to see before learning to act: Visual pre-training for manipulation,'' presented at the IEEE Int. Conf. Robot. Auto. (ICRA), Paris, France, Jun. 2020. [Online]. Available: https://yenchenlin.me/vision2action/
[76] Q. Lu, K. Chenna, B. Sundaralingam, and T. Hermans, ''Planning multi-fingered grasps as probabilistic inference in a learned deep network,'' *Robot. Res.*, vol. 10, pp. 455–472, 2020, doi: 10.1007/978-3-030-28619-4_35.
[77] A. Mousavian, C. Eppner, and D. Fox, ''6-DOF GraspNet: Variational grasp generation for object manipulation,'' in *Proc. IEEE Int. Conf. Comput. Vis.*, Seoul, South Korea, Feb. 2020, pp. 2901–2910, doi: 10.1109/ICCV.2019.00299.
[78] J. Mahler and K. Goldberg, ''Learning deep policies for robot bin picking by simulating robust grasping sequences,'' in *Proc. 1st Annu. Conf. Robot Learn. (PMLR)*, Mountain View, CA, USA, Nov. 2017, pp. 515–524.

[79] K. Wada, K. Okada, and M. Inaba, "Probabilistic 3D multilabel real-time mapping for multi-object manipulation," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Vancouver, BC, Canada, Dec. 2017, pp. 5092–5099, doi: 10.1109/IROS.2017.8206394.

[80] I. Lenz, H. Lee, and A. Saxena, "Deep learning for detecting robotic grasps," *Int. J. Robot. Res.*, vol. 34, nos. 4–5, pp. 705–724, Apr. 2015, doi: 10.1177/0278364914549607.

[81] A. Herzog, P. Pastor, M. Kalakrishnan, L. Righetti, T. Asfour, and S. Schaal, "Template-based learning of grasp selection," in *Proc. IEEE Int. Conf. Robot. Autom.*, Saint Paul, MN, USA, Jun. 2012, pp. 2379–2384, doi: 10.1109/ICRA.2012.6225271.

[82] A. Herzog, P. Pastor, M. Kalakrishnan, L. Righetti, J. Bohg, T. Asfour, and S. Schaal, "Learning of grasp selection based on shape-templates," *Auto. Robots*, vol. 36, nos. 1–2, pp. 51–65, Jan. 2014, doi: 10.1007/s10514-013-9366-8.

[83] F. J. Chu, R. Xu, and P. A. Vela, "Real-world multiobject, multigrasp detection," *IEEE Robot. Autom. Lett.*, vol. 3, no. 4, pp. 3355–3362, Oct. 2018, doi: 10.1109/LRA.2018.2852777.

[84] M. Bonilla, D. Resasco, M. Gabiccini, and A. Bicchi, "Grasp planning with soft hands using Bounding Box object decomposition," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Hamburg, Germany, Dec. 2015, pp. 518–523, doi: 10.1109/IROS.2015.7353421.

[85] P. Ni, W. Zhang, W. Bai, M. Lin, and Q. Cao, "A new approach based on two-stream CNNs for novel objects grasping in clutter," *J. Intell. Robot. Syst. Theory Appl.*, vol. 2018, pp. 1–17, Jan. 2018, doi: 10.1007/s10846-018-0788-6.

[86] D. Fischinger, M. Vincze, and Y. Jiang, "Learning grasps for unknown objects in cluttered scenes," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, Karlsruhe, Germany, Oct. 2013, pp. 609–616, doi: 10.1109/ICRA.2013.6630636.

[87] M. Gualtieri, A. ten Pas, K. Saenko, and R. Platt, "High precision grasp pose detection in dense clutter," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Daejeon, South Korea, Oct. 2016, pp. 598–605, doi: 10.1109/IROS.2016.7759114.

[88] K. Wada, M. Murooka, K. Okada, and M. Inaba, "3D object segmentation for shelf bin picking by humanoid with deep learning and occupancy voxel grid map," in *Proc. IEEE-RAS 16th Int. Conf. Hum. Robots (Humanoids)*, Cancún, Mexico, Nov. 2016, pp. 1149–1154, doi: 10.1109/HUMANOIDS.2016.7803415.

[89] A. Zeng, K.-T. Yu, S. Song, D. Suo, E. Walker, A. Rodriguez, and J. Xiao, "Multi-view self-supervised deep learning for 6D pose estimation in the Amazon picking challenge," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Singapore, May 2017, pp. 1386–1393, doi: 10.1109/ICRA.2017.7989165.

[90] D. Morrison, P. Corke, and J. Leitner, "Closing the loop for robotic grasping: A real-time, generative grasp synthesis approach," 2018, pp. 1–10, *arXiv:1804.05172*. [Online]. Available: http://arxiv.org/abs/1804.05172

[91] B. Wu, I. Akinola, A. Gupta, F. Xu, J. Varley, D. Watkins-Valls, and P. K. Allen, "Generative attention learning: A 'GenerAL' framework for high-performance multi-fingered grasping in clutter," *Auto. Robots*, vol. 44, no. 6, pp. 971–990, Jul. 2020, doi: 10.1007/s10514-020-09907-y.

[92] K. Wu, R. Ranasinghe, and G. Dissanayake, "Active recognition and pose estimation of household objects in clutter," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Seattle, WA, USA, May 2015, pp. 4230–4237, doi: 10.1109/ICRA.2015.7139782.

[93] T. Novkovic, R. Pautrat, F. Furrer, M. Breyer, R. Siegwart, and J. Nieto, "Object finding in cluttered scenes using interactive perception," 2019, pp. 1–7, *arXiv:1911.07482*. [Online]. Available: http://arxiv.org/abs/1911.07482

[94] D. Jiang, H. Wang, W. Chen, and R. Wu, "A novel occlusion-free active recognition algorithm for objects in clutter," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Qingdao, China, Mar. 2016, pp. 1389–1394, doi: 10.1109/ROBIO.2016.7866521.

[95] R. Cheng, A. Agarwal, and K. Fragkiadaki, "Reinforcement learning of active vision for manipulating objects under occlusions," in *Proc. 2nd Conf. Robot Learn. (CoRL)*. Zürich, Switzerland: PMLR, Oct. 2018, pp. 422–431.

[96] A. Zeng *et al.*, "Robotic pick-and-place of novel objects in clutter with multi-affordance grasping and cross-domain image matching," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Brisbane, QLD, Australia, Sep. 2018, pp. 3750–3757, doi: 10.1109/ICRA.2018.8461044.

[97] L. Berscheid, P. Meisner, and T. Kroger, "Self-supervised learning for precise pick-and-place without object model," *IEEE Robot. Autom. Lett.*, vol. 5, no. 3, pp. 4828–4835, Jul. 2020, doi: 10.1109/LRA.2020.3003865.

[98] T. Kim, Y. Park, Y. Park, and I. H. Suh, "Acceleration of actor-critic deep reinforcement learning for visual grasping in clutter by state representation learning based on disentanglement of a raw input image," 2020, pp. 1–8, *arXiv:2002.11903*. [Online]. Available: http://arxiv.org/abs/2002.11903

[99] M. Vohra, R. Prakash, and L. Behera, "Real-time grasp pose estimation for novel objects in densely cluttered environment," in *Proc. 28th IEEE Int. Conf. Robot Human Interact. Commun. (RO-MAN)*, New Delhi, India, Oct. 2019, pp. 1–6, doi: 10.1109/RO-MAN46459.2019.8956438.

[100] Y. Qin, R. Chen, H. Zhu, M. Song, J. Xu, and H. Su, "S4G: Amodal single-view single-shot SE(3) grasp detection in cluttered scenes," 2019, pp. 1–16, *arXiv:1910.14218*. [Online]. Available: http://arxiv.org/abs/1910.14218

[101] M. S. Kopicki, D. Belter, and J. L. Wyatt, "Learning better generative models for dexterous, single-view grasping of novel objects," *Int. J. Robot. Res.*, vol. 38, nos. 10–11, pp. 1246–1267, Sep. 2019, doi: 10.1177/0278364919865338.

[102] A. Murali, A. Mousavian, C. Eppner, C. Paxton, and D. Fox, "6-DOF grasping for target-driven object manipulation in clutter," 2019, pp. 1–7, *arXiv:1912.03628*. [Online]. Available: http://arxiv.org/abs/1912.03628

[103] Q. Shao, J. Hu, W. Wang, Y. Fang, W. Liu, J. Qi, and J. Ma, "Suction grasp region prediction using self-supervised learning for object picking in dense clutter," in *Proc. IEEE 5th Int. Conf. Mechatronics Syst. Robots (ICMSR)*, Singapore, May 2019, pp. 7–12, doi: 10.1109/ICMSR.2019.8835468.

[104] M. T. Mason, "Mechanics and planning of manipulator pushing operations," *Int. J. Robot. Res.*, vol. 5, no. 3, pp. 53–71, Sep. 1986, doi: 10.1177/027836498600500303.

[105] S. Goyal, A. Ruina, and J. Papadopoulos, "Planar sliding with dry friction Part 1. Limit surface and moment function," *Wear*, vol. 143, no. 2, pp. 307–330, 1991, doi: 10.1016/0043-1648(91)90104-3.

[106] K.-T. Yu, M. Bauza, N. Fazeli, and A. Rodriguez, "More than a million ways to be pushed. A high-fidelity experimental dataset of planar pushing," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Daejeon, South Korea, Dec. 2016, pp. 30–37, doi: 10.1109/IROS.2016.7758091.

[107] M. Bauza and A. Rodriguez, "A probabilistic data-driven model for planar pushing," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Singapore, May 2017, pp. 3008–3015, doi: 10.1109/ICRA.2017.7989345.

[108] C. Finn, I. Goodfellow, and S. Levine, "Unsupervised learning for physical interaction through video prediction," in *Proc. 30th Int. Conf. Neural Inf. Process. Syst. (NIPS)*, Red Hook, NY, USA, Dec. 2016, pp. 64–72, doi: 10.5555/3157096.3157104.

[109] C. Finn and S. Levine, "Deep visual foresight for planning robot motion," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Singapore, Jul. 2017, pp. 2786–2793, doi: 10.1109/ICRA.2017.7989324.

[110] J. Stüber, C. Zito, and R. Stolkin, "Let's push things forward: A survey on robot pushing," *Frontiers Robot. AI*, vol. 7, p. 8, Feb. 2020, doi: 10.3389/frobt.2020.00008.

[111] S. James, A. J. Davison, and E. Johns, "Transferring end-to-end visuomotor control from simulation to real world for a multi-stage task," in *Proc. 1st Conf. Robot Learn. (CoRL)*, Mountain View, CA, USA, Nov. 2017, pp. 1–10.

[112] S. Levine, P. Pastor, A. Krizhevsky, J. Ibarz, and D. Quillen, "Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection," *Int. J. Robot. Res.*, vol. 37, nos. 4–5, pp. 421–436, Apr. 2018, doi: 10.1177/0278364917710318.

[113] F. Ebert, C. Finn, A. X. Lee, and S. Levine, "Self-supervised visual planning with temporal skip connections," in *Proc. 1st Conf. Robot Learn. (CoRL)*, Mountain View, CA, USA, Nov. 2017, pp. 1–13.

[114] W. Bejjani, R. Papallas, M. Leonetti, and M. R. Dogar, "Planning with a receding horizon for manipulation in clutter using a learned value function," in *Proc. IEEE-RAS 18th Int. Conf. Hum. Robots (Humanoids)*, Beijing, China, Nov. 2018, pp. 1–9, doi: 10.1109/HUMANOIDS.2018.8624977.

[115] W. Bejjani, M. R. Dogar, and M. Leonetti, "Learning physics-based manipulation in clutter: Combining image-based generalization and look-ahead planning," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Macau, China, Nov. 2019, pp. 6562–6569.

[116] W. C. Agboh and M. R. Dogar, "Real-time online re-planning for grasping under clutter and uncertainty," in *Proc. IEEE-RAS 18th Int. Conf. Hum. Robots (Humanoids)*, Beijing, China, Nov. 2018, pp. 387–394, doi: 10.1109/HUMANOIDS.2018.8625041.

[117] K. Lowrey, A. Rajeswaran, S. Kakade, E. Todorov, and I. Mordatch, "Plan online, learn offline: Efficient learning and exploration via model-based control," 2018, pp. 1–15, *arXiv:1811.01848*. [Online]. Available: http://arxiv.org/abs/1811.01848

[118] H. Song, J. A. Haustein, W. Yuan, K. Hang, M. Yu Wang, D. Kragic, and J. A. Stork, "Multi-object rearrangement with Monte Carlo tree search: A case study on planar nonprehensile sorting," 2019, pp. 1–8, *arXiv:1912.07024*. [Online]. Available: http://arxiv.org/abs/1912.07024

[119] A. Zeng, S. Song, S. Welker, J. Lee, A. Rodriguez, and T. Funkhouser, "Learning synergies between pushing and grasping with self-supervised deep reinforcement learning," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Madrid, Spain, Jan. 2018, pp. 4238–4245, doi: 10.1109/IROS.2018.8593986.

[120] Y. Yang, H. Liang, and C. Choi, "A deep learning approach to grasping the invisible," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 2232–2239, Apr. 2020, doi: 10.1109/LRA.2020.2970622.

[121] I. Sarantopoulos, M. Kiatos, Z. Doulgeri, and S. Malassiotis, "Split deep Q-learning for robust object singulation," 2019, pp. 1–7, *arXiv:1909.08105*. [Online]. Available: http://arxiv.org/abs/1909.08105

[122] A. Hundt, B. Killeen, N. Greene, H. Wu, H. Kwon, C. Paxton, and G. D. Hager, "'Good robot!': Efficient reinforcement learning for multi-step visual tasks with sim to real transfer," 2019, pp. 1–8, *arXiv:1909.11730*. [Online]. Available: http://arxiv.org/abs/1909.11730

[123] L. Berscheid, P. Meifbner, and T. Kroger, "Robot learning of shifting objects for grasping in cluttered environments," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Macau, China, Jan. 2020, pp. 612–618, 2020, doi: 10.1109/IROS40897.2019.8968042.

[124] M. Kiatos and S. Malassiotis, "Robust object grasping in clutter via singulation," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, Montreal, QC, Canada, Aug. 2019, pp. 1596–1600, doi: 10.1109/ICRA.2019.8793972.

[125] T. Silver, K. Allen, J. Tenenbaum, and L. Kaelbling, "Residual policy learning," 2019, pp. 1–6, *arXiv:1812.06298*. [Online]. Available: https://arxiv.org/abs/1812.06298

[126] P. Ni, W. Zhang, H. Zhang, and Q. Cao, "Learning efficient push and grasp policy in a totebox from simulation," *Adv. Robot.*, vol. 34, no. 13, pp. 1–15, May 2020, doi: 10.1080/01691864.2020.1757504.

[127] Z. Sun, K. Yuan, W. Hu, C. Yang, and Z. Li, "Learning pregrasp manipulation of objects from ungraspable poses," 2020, pp. 1–8, *arXiv:2002.06344*. [Online]. Available: https://arxiv.org/abs/2002.06344

[128] H. J. Jo and J. B. Song, "Object-independent grasping in heavy clutter," *Appl. Sci.*, vol. 10, no. 3, pp. 804–894, Jan. 2020, doi: 10.3390/app10030804.

[129] H. Liu, Y. Yuan, Y. Deng, X. Guo, Y. Wei, K. Lu, B. Fang, D. Guo, and F. Sun, "Active affordance exploration for robot grasping," in *Proc. 12th Int. Conf. Intell. Robot. Appl. (ICIRA)*. Shenyang, China: Springer, Aug. 2019, pp. 426–438, doi: 10.1007/978-3-030-27541-9_35.

[130] M. Danielczuk, A. Kurenkov, A. Balakrishna, M. Matl, D. Wang, R. Martin-Martin, A. Garg, S. Savarese, and K. Goldberg, "Mechanical search: Multi-step retrieval of a target object occluded by clutter," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, Montreal, QC, Canada, Aug. 2019, pp. 1614–1621, doi: 10.1109/ICRA.2019.8794143.

[131] M. Yan, A. Li, M. Kalakrishnan, and P. Pastor, "Learning probabilistic multi-modal actor models for vision-based robotic grasping," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, Montreal, QC, Canada, Aug. 2019, pp. 4804–4810, doi: 10.1109/ICRA.2019.8794024.

[132] K. Wada, K. Okada, and M. Inaba, "Joint learning of instance and semantic segmentation for robotic Pick-and-Place with heavy occlusions in clutter," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, May 2019, pp. 9558–9564, doi: 10.1109/ICRA.2019.8793783.

[133] Y. Song, Y. Fei, C. Cheng, X. Li, and C. Yu, "UG-net for robotic grasping using only depth image," in *Proc. IEEE Int. Conf. Real-Time Comput. Robot. (RCAR)*, Irkutsk, Russia, Mar. 2019, pp. 913–918, doi: 10.1109/RCAR47638.2019.9044116.

[134] W. Guo, G. Dong, C. Chen, and M. Li, "Learning pushing skills using object detection and deep reinforcement learning," in *Proc. IEEE Int. Conf. Mechatronics Autom. (ICMA)*, Tianjin, China, Aug. 2019, pp. 469–474, doi: 10.1109/ICMA.2019.8816481.

[135] M. Han, W. Liu, Z. Pan, T. Xue, Q. Shao, J. Ma, and W. Wang, "Object-agnostic suction grasp affordance detection in dense cluster using self-supervised learning.Docx," 2019, pp. 1–6, *arXiv:1906.02995*. [Online]. Available: http://arxiv.org/abs/1906.02995

[136] C. Chen, H.-Y. Li, X. Zhang, X. Liu, and U.-X. Tan, "Towards robotic picking of targets with background distractors using deep reinforcement learning," in *Proc. WRC Symp. Adv. Robot. Autom. (WRC SARA)*, Beijing, China, Aug. 2019, pp. 166–171, doi: 10.1109/WRC-SARA.2019.8931932.

[137] W. Yuan, J. A. Stork, D. Kragic, M. Y. Wang, and K. Hang, "Rearrangement with nonprehensile manipulation using deep reinforcement learning," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Brisbane, QLD, Australia, May 2018, pp. 270–277, doi: 10.1109/ICRA.2018.8462863.

[138] Y. Labbe, S. Zagoruyko, I. Kalevatykh, I. Laptev, J. Carpentier, M. Aubry, and J. Sivic, "Monte-carlo tree search for efficient visually guided rearrangement planning," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 3715–3722, Apr. 2020, doi: 10.1109/LRA.2020.2980984.

[139] S. H. Cheong, B. Y. Cho, J. Lee, C. Kim, and C. Nam, "Where to relocate: Object rearrangement inside cluttered and confined environments for robotic manipulation," presented at the IEEE Int. Conf. Robot. Auto. (ICRA), Paris, France, Jun. 2020. [Online]. Available: http://arxiv.org/abs/2003.10863

[140] E. Huang, Z. Jia, and M. T. Mason, "Large-scale multi-object rearrangement," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, Montreal, QC, Canada, Aug. 2019, pp. 211–218, doi: 10.1109/ICRA.2019.8793946.

[141] H. van Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double Q-learning," in *Proc. 13th AAAI Conf. Artif. Intell. (AAAI)*, Phoenix, AZ, USA, Feb. 2016, pp. 2094–2100. [Online]. Available: http://arxiv.org/abs/1509.06461

[142] Z. Wang, T. Schaul, M. Hessel, H. van Hasselt, M. Lanctot, and N. de Freitas, "Dueling network architectures for deep reinforcement learning," 2015, pp. 2939–2947, *arXiv:1511.06581*. [Online]. Available: http://arxiv.org/abs/1511.06581

[143] O. Kroemer and G. S. Sukhatme, "Feature selection for learning versatile manipulation skills based on observed and desired trajectories," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Singapore, Jul. 2017, pp. 4713–4720, doi: 10.1109/ICRA.2017.7989546.

[144] U. Viereck, A. ten Pas, K. Saenko, and R. Platt, "Learning a visuomotor controller for real world robotic grasping using simulated depth images," in *Proc. 1st Conf. Robot Learn. (CoRL)*, Mountain View, CA, USA, Nov. 2017, pp. 1–10. [Online]. Available: http://arxiv.org/abs/1706.04652

[145] A. A. Rusu, M. Vecerik, T. Rothörl, N. Heess, R. Pascanu, and R. Hadsell, "Sim-to-real robot learning from pixels with progressive nets," 2016, pp. 1–9, *arXiv:1610.04286*. [Online]. Available: http://arxiv.org/abs/1610.04286

[146] A. A. Rusu, N. C. Rabinowitz, G. Desjardins, H. Soyer, J. Kirkpatrick, K. Kavukcuoglu, R. Pascanu, and R. Hadsell, "Progressive neural networks," 2016, pp. 1–14, *arXiv:1606.04671*. [Online]. Available: http://arxiv.org/abs/1606.04671

[147] P. Christiano, Z. Shah, I. Mordatch, J. Schneider, T. Blackwell, J. Tobin, P. Abbeel, and W. Zaremba, "Transfer from simulation to real world through learning deep inverse dynamics model," 2016, pp. 1–8, *arXiv:1610.03518*. [Online]. Available: http://arxiv.org/abs/1610.03518

[148] J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, and P. Abbeel, "Domain randomization for transferring deep neural networks from simulation to the real world," 2017, pp. 1–8, *arXiv:1703.06907*. [Online]. Available: http://arxiv.org/abs/1703.06907

[149] F. Zhang, J. Leitner, M. Milford, and P. Corke, "Modular deep Q networks for sim-to-real transfer of visuo-motor policies," in *Proc. Australas. Conf. Robot. Autom. (ACRA)*. Ultimo, NSW, Australia: Univ. of Technology Sydney, Dec. 2017, pp. 62–71. [Online]. Available: http://arxiv.org/abs/1610.06781

[150] L. Pinto, M. Andrychowicz, P. Welinder, W. Zaremba, and P. Abbeel, "Asymmetric actor critic for image-based robot learning," 2017, pp. 1–8, *arXiv:1710.06542*. [Online]. Available: http://arxiv.org/abs/1710.06542

[151] J. Matas, S. James, and A. J. Davison, "Sim-to-real reinforcement learning for deformable object manipulation," in *Proc. 2nd Conf. Robot Learn. (CoRL)*, Zürich, Switzerland, Oct. 2018, pp. 1–8. [Online]. Available: http://hdl.handle.net/10044/1/64366

[152] O. M. Andrychowicz, B. Baker, M. Chociej, R. Józefowicz, B. McGrew, J. Pachocki, A. Petron, M. Plappert, G. Powell, A. Ray, J. Schneider, S. Sidor, J. Tobin, P. Welinder, L. Weng, and W. Zaremba, "Learning dexterous in-hand manipulation," *Int. J. Robot. Res.*, vol. 39, no. 1, pp. 3–20, Jan. 2020, doi: 10.1177/0278364919887447.

[153] F. Sadeghi, A. Toshev, E. Jang, and S. Levine, "Sim2Real viewpoint invariant visual servoing by recurrent control," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Dec. 2018, pp. 4691–4699, doi: 10.1109/CVPR.2018.00493.
[154] Y. Chebotar, A. Handa, V. Makoviychuk, M. Macklin, J. Issac, N. Ratliff, and D. Fox, "Closing the sim-to-real Loop: Adapting simulation randomization with real world experience," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Montreal, QC, Canada, Aug. 2019, pp. 8973–8979, doi: 10.1109/ICRA.2019.8793789.
[155] S. James, P. Wohlhart, M. Kalakrishnan, D. Kalashnikov, A. Irpan, J. Ibarz, S. Levine, R. Hadsell, and K. Bousmalis, "Sim-to-real via sim-to-sim: Data-efficient robotic grasping via randomized-to-canonical adaptation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 12619–12629, doi: 10.1109/CVPR.2019.01291.
[156] A. Barreto, D. Borsa, J. Quan, T. Schaul, D. Silver, M. Hessel, D. Mankowitz, A. Žídek, and R. Munos, "Transfer in deep reinforcement learning using successor features and generalised policy improvement," in *Proc. 35th Int. Conf. Mach. Learn. (ICML)*, Stockholm, Sweden, Jul. 2018, pp. 844–853. [Online]. Available: http://arxiv.org/abs/1901.10964
[157] A. Amiranashvili, A. Dosovitskiy, V. Koltun, and T. Brox, "Motion perception in reinforcement learning with dynamic objects," in *Proc. 2nd Conf. Robot Learn. (CoRL)*, Zrich, Switzerland, Oct. 2018, pp. 1–14. [Online]. Available: http://arxiv.org/abs/1901.03162
[158] X. B. Peng, M. Andrychowicz, W. Zaremba, and P. Abbeel, "Sim-to-real transfer of robotic control with dynamics randomization," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Brisbane, QLD, Australia, Sep. 2018, pp. 3803–3810, doi: 10.1109/ICRA.2018.8460528.
[159] I. Clavera, D. Held, and P. Abbeel, "Policy transfer via modularity and reward guiding," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Vancouver, BC, Canada, Dec. 2017, pp. 1537–1544, doi: 10.1109/IROS.2017.8205959.
[160] P. Kurrek, M. Jocas, F. Zoghlami, M. Stoelen, and V. Salehi, "AI motion control—A generic approach to develop control policies for robotic manipulation tasks," in *Proc. Int. Conf. Eng. Design (ICED)*, Delft, The Netherlands, Aug. 2019, pp. 3561–3570, doi: 10.1017/dsi.2019.363.
[161] M. Matthias and B. Ghanem, "Driving policy transfer via modularity and abstraction," in *Proc. 2nd Conf. Robot Learn. (PMLR)*, Zrich, Switzerland, vol. 87, Oct. 2018, pp. 1–15. [Online]. Available: http://arxiv.org/abs/1804.09364
[162] W. Yuan, K. Hang, H. Song, D. Kragic, M. Y. Wang, and J. A. Stork, "Reinforcement learning in topology-based representation for human body movement with whole arm manipulation," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, Montreal, QC, Canada, Aug. 2019, pp. 2153–2160, doi: 10.1109/ICRA.2019.8794160.
[163] W. Yuan, K. Hang, D. Kragic, M. Y. Wang, and J. A. Stork, "End-to-end nonprehensile rearrangement with deep reinforcement learning and simulation-to-reality transfer," *Robot. Auton. Syst.*, vol. 119, pp. 119–134, Sep. 2019, doi: 10.1016/j.robot.2019.06.007.
[164] J. Yang, B. Petersen, H. Zha, and D. Faissol, "Single episode policy transfer in reinforcement learning," presented at the Int. Conf. Learn. Represent. (ICLR), Formerly Addis Ababa, Ethiopia, May 2020. [Online]. Available: http://arxiv.org/abs/1910.07719
[165] W. Yu, J. Tan, C. K. Liu, and G. Turk, "Preparing for the unknown: Learning a universal policy with online system identification," 2017, pp. 1–10, *arXiv:1702.02453*. [Online]. Available: http://arxiv.org/abs/1702.02453
[166] I. Arnekvist, D. Kragic, and J. A. Stork, "VPE: Variational policy embedding for transfer reinforcement learning," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, Montreal, QC, Canada, Aug. 2019, pp. 36–42, doi: 10.1109/ICRA.2019.8793556.
[167] A. Pashevich, R. Strudel, I. Kalevatykh, I. Laptev, and C. Schmid, "Learning to augment synthetic images for Sim2Real policy transfer," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Macau, China, Nov. 2019, pp. 2651–2657, doi: 10.1109/IROS40897.2019.8967622.
[168] K. Bousmalis, A. Irpan, P. Wohlhart, Y. Bai, M. Kelcey, M. Kalakrishnan, L. Downs, J. Ibarz, P. Pastor, K. Konolige, S. Levine, and V. Vanhoucke, "Using simulation and domain adaptation to improve efficiency of deep robotic grasping," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Brisbane, QLD, Australia, May 2018, pp. 4243–4250.
[169] O. Pedersen, E. Misimi, and F. Chaumette, "Grasping unknown objects by coupling deep reinforcement learning, generative adversarial networks, and visual servoing," presented at the IEEE Int. Conf. Robot. Auto. (ICRA), Paris, France, Jun. 2020. [Online]. Available: https://hal.inria.fr/hal-02495837
[170] F. Zhang, J. Leitner, Z. Ge, M. Milford, and P. Corke, "Adversarial discriminative sim-to-real transfer of visuo-motor policies," *Int. J. Robot. Res.*, vol. 38, nos. 10–11, pp. 1229–1245, Aug. 2019, doi: 10.1177/0278364919870227.
[171] T. Dai, K. Arulkumaran, T. Gerbert, S. Tukra, F. Behbahani, and A. A. Bharath, "Analysing deep reinforcement learning agents trained with domain randomisation," 2019, pp. 1–23, *arXiv:1912.08324*. [Online]. Available: http://arxiv.org/abs/1912.08324
[172] B. Wu, I. Akinola, J. Varley, and P. Allen, "MAT: Multi-fingered adaptive tactile grasping via deep reinforcement learning," in *Proc. Conf. Robot Learn. (CoRL)*, Osaka, Japan, Oct./Nov. 2019, pp. 142–161. [Online]. Available: http://arxiv.org/abs/1909.04787
[173] J. Kober and J. Peter, "Policy search for motor primitives in robotics," *Mach. Learn.*, vol. 84, nos. 1–2, pp. 171–203, Jul. 2011, doi: 10.1007/s10994-010-5223-6.
[174] J. Peters and S. Schaal, "Reinforcement learning of motor skills with policy gradients," *Neural Netw.*, vol. 21, no. 4, pp. 682–697, May 2008, doi: 10.1016/j.neunet.2008.02.003.
[175] C. Finn, S. Levine, and P. Abbeel, "Guided cost learning: Deep inverse optimal control via policy optimization," in *Proc. 33rd Int. Conf. Mach. Learn. (ICML)*, New York, NY, USA, Jun. 2016, pp. 49–58. [Online]. Available: http://arxiv.org/abs/1603.00448
[176] M. Wulfmeier, P. Ondruska, and I. Posner, "Maximum entropy deep inverse reinforcement learning," 2015, pp. 1–8, *arXiv:1507.04888*. [Online]. Available: http://arxiv.org/abs/1507.04888
[177] T. Hester, M. Vecerik, O. Pietquin, M. Lanctot, T. Schaul, B. Piot, D. Horgan, J. Quan, A. Sendonaris, G. Dulac-Arnold, I. Osband, J. Agapiou, J. Z. Leibo, and A. Gruslys, "Deep Q-learning from Demonstrations," presented at the AAAI, Nov. 2017. [Online]. Available: http://arxiv.org/abs/1704.03732
[178] A. Nair, B. McGrew, M. Andrychowicz, W. Zaremba, and P. Abbeel, "Overcoming exploration in reinforcement learning with demonstrations," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Brisbane, QLD, Australia, Sep. 2018, pp. 6292–6299, doi: 10.1109/ICRA.2018.8463162.
[179] A. Rajeswaran, V. Kumar, A. Gupta, G. Vezzani, J. Schulman, E. Todorov, and S. Levine, "Learning complex dexterous manipulation with deep reinforcement learning and demonstrations," 2017, pp. 1–9, *arXiv:1709.10087*. [Online]. Available: http://arxiv.org/abs/1709.10087
[180] H. Zhu, A. Gupta, A. Rajeswaran, S. Levine, and V. Kumar, "Dexterous manipulation with deep reinforcement learning: Efficient, general, and low-cost," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, Montreal, QC, Canada, May 2019, pp. 3651–3657, doi: 10.1109/ICRA.2019.8794102.
[181] A. M. Ghalamzan E, C. Paxton, G. D. Hager, and L. Bascetta, "An incremental approach to learning generalizable robot tasks from human demonstration," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Seattle, WA, USA, May 2015, pp. 5616–5621, doi: 10.1109/ICRA.2015.7139985.
[182] N. Ratliff, M. Toussaint, and S. Schaal, "Understanding the geometry of workspace obstacles in motion optimization," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Seattle, WA, USA, Jul. 2015, pp. 4202–4209, doi: 10.1109/ICRA.2015.7139778.
[183] A. Rai, G. Sutanto, S. Schaal, and F. Meier, "Learning feedback terms for reactive planning and control," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Singapore, Jul. 2017, pp. 2184–2191, doi: 10.1109/ICRA.2017.7989252.
[184] A. M. Ghalamzan E. and M. Ragaglia, "Robot learning from demonstrations: Emulation learning in environments with moving obstacles," *Robot. Auto. Syst.*, vol. 101, pp. 45–56, Mar. 2018, doi: 10.1016/j.robot.2017.12.001.
[185] A. Hussein, M. M. Gaber, E. Elyan, and C. Jayne, "Imitation learning: A survey of learning methods," *ACM Comput. Surv.*, vol. 50, no. 2, pp. 1–35, Jun. 2017, doi: 10.1145/3054912.
[186] R. Li, R. Platt, W. Yuan, A. ten Pas, N. Roscup, M. A. Srinivasan, and E. Adelson, "Localization and manipulation of small parts using GelSight tactile sensing," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Chicago, IL, USA, Nov. 2014, pp. 3988–3993, doi: 10.1109/IROS.2014.6943123.
[187] A. Tamar, G. Thomas, T. Zhang, S. Levine, and P. Abbeel, "Learning from the hindsight plan—Episodic MPC improvement," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Singapore, Jul. 2017, pp. 336–343, doi: 10.1109/ICRA.2017.7989043.
[188] T. Inoue, G. De Magistris, A. Munawar, T. Yokoya, and R. Tachibana, "Deep reinforcement learning for high precision assembly tasks," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Vancouver, BC, Canada, Dec. 2017, pp. 819–825, doi: 10.1109/IROS.2017.8202244.

[189] J. Luo, E. Solowjow, C. Wen, J. A. Ojea, A. M. Agogino, A. Tamar, and P. Abbeel, "Reinforcement learning on variable impedance controller for high-precision robotic assembly," in *Proc. Int. Conf. Robot. And Autom. (ICRA)*, Montreal, QC, Canada, Aug. 2019, pp. 3080–3087, doi: 10.1109/ICRA.2019.8793506.
[190] G. Schoettler, A. Nair, J. Luo, S. Bahl, J. A. Ojea, E. Solowjow, and S. Levine, "Deep reinforcement learning for industrial insertion tasks with visual inputs and natural rewards," 2019, pp. 1–10, *arXiv:1906.05841*. [Online]. Available: http://arxiv.org/abs/1906.05841
[191] Y. Zhu, Z. Wang, J. Merel, A. Rusu, T. Erez, S. Cabi, S. Tunyasuvunakool, J. Kramár, R. Hadsell, N. de Freitas, and N. Heess, "Reinforcement and imitation learning for diverse visuomotor skills," 2018, pp. 1–12, *arXiv:1802.09564*. [Online]. Available: http://arxiv.org/abs/1802.09564
[192] M. Chevalier-Boisvert, G. Alain, F. Golemo, and D. Nowrouzezahrai, "Robo-PlaNet: Learning to poke in a day," 2019, pp. 1–6, *arXiv:1911.03594*. [Online]. Available: http://arxiv.org/abs/1911.03594
[193] T. Johannink, S. Bahl, A. Nair, J. Luo, A. Kumar, M. Loskyll, J. A. Ojea, E. Solowjow, and S. Levine, "Residual reinforcement learning for robot control," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, Montreal, QC, Canada, Aug. 2019, pp. 6023–6029, doi: 10.1109/ICRA.2019.8794127.
[194] G. Liu, C. Song, X. Zang, and J. Zhao, "Reactive execution of learned tasks with real-time collision avoidance in a dynamic environment," *IEEE Access*, vol. 6, pp. 57366–57375, Oct. 2018, doi: 10.1109/ACCESS.2018.2873718.
[195] Z. Zhu and H. Hu, "Robot learning from demonstration in robotic assembly: A survey," *Robotics*, vol. 7, no. 2, p. 17, Apr. 2018, doi: 10.3390/robotics7020017.
[196] B. Fang, S. Jia, D. Guo, M. Xu, S. Wen, and F. Sun, "Survey of imitation learning for robotic manipulation," *Int. J. Intell. Robot. Appl.*, vol. 3, no. 4, pp. 362–369, Sep. 2019, doi: 10.1007/s41315-019-00103-5.
[197] S. Piao, Y. Huang, and H. Liu, "Online multi-modal imitation learning via lifelong intention encoding," in *Proc. IEEE 4th Int. Conf. Adv. Robot. Mechatronics (ICARM)*, Toyonaka, Japan, Jul. 2019, pp. 786–792, doi: 10.1109/ICARM.2019.8833960.
[198] D. Seita, N. Jamali, M. Laskey, A. K. Tanwani, R. Berenstein, P. Baskaran, S. Iba, J. Canny, and K. Goldberg, "Deep transfer learning of pick points on fabric for robot bed-making," 2018, pp. 1–16, *arXiv:1809.09810*. [Online]. Available: http://arxiv.org/abs/1809.09810
[199] Y. Cui, D. Isele, S. Niekum, and K. Fujimura, "Uncertainty-aware data aggregation for deep imitation learning," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, Montreal, QC, Canada, Aug. 2019, pp. 761–767, doi: 10.1109/ICRA.2019.8794025.
[200] S. Arnold and K. Yamazaki, "Fast and flexible multi-step cloth manipulation planning using an encode-manipulate-decode network (EM*D Net)," *Frontiers Neurorobot.*, vol. 13, p. 22, May 2019, doi: 10.3389/fnbot.2019.00022.
[201] R. Jangir, G. Alenya, and C. Torras, "Dynamic cloth manipulation with deep reinforcement learning," presented at the Int. Conf. Robot. Auto. (ICRA), Paris, France, Jun. 2020. [Online]. Available: http://arxiv.org/abs/1910.14475
[202] A. Cherubini, V. Ortenzi, A. Cosgun, R. Lee, and P. Corke, "Model-free vision-based shaping of deformable plastic materials," *Int. J. Robot. Res.*, pp. 1–21, Feb. 2020, doi: 10.1177/0278364920907684.
[203] A. Verleysen, M. Biondina, and F. Wyffels, "Video dataset of human demonstrations of folding clothing for robotic folding," *Int. J. Robot. Res.*, vol. 39, no. 9, pp. 1031–1036, Jul. 2020, doi: 10.1177/0278364920940408.
[204] P. Abolghasemi and L. Bölöni, "Accept synthetic objects as real: End-to-end training of attentive deep visuomotor policies for manipulation in clutter," 2019, pp. 1–7, *arXiv:1909.11128*. [Online]. Available: http://arxiv.org/abs/1909.11128
[205] S. Song, A. Zeng, J. Lee, and T. Funkhouser, "Grasping in the wild: Learning 6DoF closed-loop grasping from low-cost demonstrations," *IEEE Robot. Autom. Lett.*, vol. 5, no. 3, pp. 4978–4985, Jul. 2020, doi: 10.1109/LRA.2020.3004787.
[206] D. Kappler, J. Bohg, and S. Schaal, "Leveraging big data for grasp planning," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Seattle, WA, USA, Jul. 2015, pp. 4304–4311, doi: 10.1109/ICRA.2015.7139793.
[207] J. Mahler, F. T. Pokorny, B. Hou, M. Roderick, M. Laskey, M. Aubry, K. Kohlhoff, T. Kröger, J. Kuffner, and K. Goldberg, "Dex-Net 1.0: A cloud-based network of 3D objects for robust grasp planning using a multi-armed bandit model with correlated rewards," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Stockholm, Sweden, Jun. 2016, pp. 1957–1964, doi: 10.1109/ICRA.2016.7487342.
[208] 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," 2017, pp. 1–12, *arXiv:1703.09312*. [Online]. Available: http://arxiv.org/abs/1703.09312
[209] J. Mahler, M. Matl, X. Liu, A. Li, D. Gealy, and K. Goldberg, "Dex-net 3.0: Computing robust vacuum suction grasp targets in point clouds using a new analytic model and deep learning," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Brisbane, QLD, Australia, Sep. 2018, pp. 5620–5627, doi: 10.1109/ICRA.2018.8460887.
[210] N. Jakobi, P. Husbands, and I. Harvey, "Noise and the reality gap: The use of simulation in evolutionary robotics," in *Advances in Artificial Life* (Lecture Notes in Computer Science: Lecture Notes in Artificial Intelligence), vol. 929, F. Morán, A. Moreno, J. J. Merelo, and P. Chacón, Eds. Berlin, Germany: Springer, 1995, pp. 704–720, doi: 10.1007/3-540-59496-5_337.
[211] J. Mahler, M. Matl, V. Satish, M. Danielczuk, B. DeRose, S. McKinley, and K. Goldberg, "Learning ambidextrous robot grasping policies," *Sci. Robot.*, vol. 4, no. 26, Jan. 2019, Art. no. eaau4984, doi: 10.1126/scirobotics.aau4984.
[212] B. Staub, A. K. Tanwani, J. Mahler, M. Breyer, M. Laskey, Y. Takaoka, M. Bajracharya, R. Siegwart, and K. Goldberg, "Dex-net MM: Deep grasping for surface decluttering with a low-precision mobile manipulator," in *Proc. IEEE 15th Int. Conf. Autom. Sci. Eng. (CASE)*, Vancouver, BC, Canada, Sep. 2019, pp. 1373–1379, doi: 10.1109/COASE.2019.8842901.
[213] J. Fu, S. Levine, and P. Abbeel, "One-shot learning of manipulation skills with online dynamics adaptation and neural network priors," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Daejeon, South Korea, Dec. 2016, pp. 4019–4026, doi: 10.1109/IROS.2016.7759592.
[214] M. T. Gillespie, C. M. Best, E. C. Townsend, D. Wingate, and M. D. Killpack, "Learning nonlinear dynamic models of soft robots for model predictive control with neural networks," in *Proc. IEEE Int. Conf. Soft Robot. (RoboSoft)*, Livorno, Italy, Jul. 2018, pp. 39–45, doi: 10.1109/ROBOSOFT.2018.8404894.
[215] K. Chua, R. Calandra, R. McAllister, and S. Levine, "Deep reinforcement learning in a handful of trials using probabilistic dynamics models," in *Proc. Adv. Neur. Inf. Proc. Syst. (NIPS)*, vol. 31, 2018, pp. 4754–4765.
[216] S. Nair and C. Finn, "Hierarchical foresight: Self-supervised learning of long-horizon tasks via visual subgoal generation," 2019, pp. 1–16, *arXiv:1909.05829*. [Online]. Available: http://arxiv.org/abs/1909.05829
[217] D. Morrison, P. Corke, and J. Leitner, "Learning robust, real-time, reactive robotic grasping," *Int. J. Robot. Res.*, vol. 39, nos. 2–3, pp. 183–201, Mar. 2020, doi: 10.1177/0278364919859066.
[218] A. ten Pas, M. Gualtieri, K. Saenko, and R. Platt, "Grasp pose detection in point clouds," *Int. J. Robot. Res.*, vol. 36, nos. 13–14, pp. 1455–1473, Oct. 2017, doi: 10.1177/0278364917735594.
[219] M. Gualtieri and R. Platt, "Viewpoint selection for grasp detection," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Vancouver, BC, Canada, Dec. 2017, pp. 258–264, doi: 10.1109/IROS.2017.8202166.
[220] V. Satish, J. Mahler, and K. Goldberg, "On-policy dataset synthesis for learning robot grasping policies using fully convolutional deep networks," *IEEE Robot. Autom. Lett.*, vol. 4, no. 2, pp. 1357–1364, Apr. 2019, doi: 10.1109/LRA.2019.2895878.
[221] E. Johns, S. Leutenegger, and A. J. Davison, "Deep learning a grasp function for grasping under gripper pose uncertainty," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Daejeon, South Korea, Dec. 2016, pp. 4461–4468, doi: 10.1109/IROS.2016.7759657.
[222] X. Yan, J. Hsu, M. Khansari, Y. Bai, A. Pathak, A. Gupta, J. Davi, and H. Lee, "Learning 6-DOF grasping interaction via deep geometry-aware 3D representations," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Brisbane, QLD, Australia, Sep. 2018, pp. 1–9, doi: 10.1109/ICRA.2018.8460609.
[223] M. Liu, Z. Pan, K. Xu, K. Ganguly, and D. Manocha, "Generating grasp poses for a high-DOF gripper using neural networks," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Macau, China, Jan. 2020, pp. 1518–1525, doi: 10.1109/IROS40897.2019.8968115.
[224] C. Eppner, A. Mousavian, and D. Fox, "A billion ways to grasp: An evaluation of grasp sampling schemes on a dense, physics-based grasp data set," presented at the Int. Symp. Robot. Res. (ISRR), Hanoi, Vietnam, Oct. 2019. [Online]. Available: https://arxiv.org/abs/1912.05604
[225] A. Kalra, V. Taamazyan, S. K. Rao, K. Venkataraman, R. Raskar, and A. Kadambi, "Deep polarization cues for transparent object segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, WA, USA, Aug. 2020, pp. 8602–8611, doi: 10.1109/CVPR42600.2020.00863.

[226] X. Liu, R. Jonschkowski, A. Angelova, and K. Konolige, "KeyPose: Multi-view 3D labeling and keypoint estimation for transparent objects," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, WA, USA, Aug. 2020, pp. 11602–11604, doi: 10.1109/CVPR42600.2020.01162.
[227] T. Weng, A. Pallankize, Y. Tang, O. Kroemer, and D. Held, "Multi-modal transfer learning for grasping transparent and specular objects," *IEEE Robot. Autom. Lett.*, vol. 5, no. 3, pp. 3796–3803, Jul. 2020, doi: 10.1109/LRA.2020.2974686.
[228] S. S. Sajjan, M. Moore, M. Pan, G. Nagaraja, J. Lee, A. Zeng, and S. Song, "ClearGrasp: 3D Shape Estimation of Transparent Objects for Manipulation," presented at the Int. Conf. Robot. Auto. (ICRA), Paris, France,Jun. 2020. [Online]. Available: https://arxiv.org/abs/1910.02550
[229] P. Wu, W. Chen, H. Liu, Y. Duan, N. Lin, and X. Chen, "Predicting grasping order in clutter environment by using both color image and points cloud," in *Proc. WRC Symp. Adv. Robot. Autom. (WRC SARA)*, Beijing, China, Dec. 2019, pp. 197–202, doi: 10.1109/WRC-SARA.2019.8931929.
[230] Y. Lin, C. Tang, F.-J. Chu, and P. A. Vela, "Using synthetic data and deep networks to recognize primitive shapes for object grasping," Sep. 2019, vol. 1, pp. 1–8, *arXiv:1909.08508*. [Online]. Available: https://arxiv.org/abs/1909.08508
[231] K. Sanders, M. Danielczuk, J. Mahler, A. Tanwani, and K. Goldberg, "Non-Markov policies to reduce sequential failures in robot bin picking," Sep. 2019, vol. 1, pp. 1–8, *arXiv:2007.10420*. [Online]. Available: http://arxiv.org/abs/2007.10420
[232] P. Ni, W. Zhang, X. Zhu, and Q. Cao, "PointNet++ grasping: Learning an end-to-end spatial grasp generation algorithm from sparse point clouds," presented at the Int. Conf. Robot. Auto. (ICRA), Paris, France, Jun. 2020. [Online]. Available: http://arxiv.org/abs/2003.09644
[233] X. Chen, Z. Ye, J. Sun, Y. Fan, F. Hu, C. Wang, and C. Lu, "Transferable active grasping and real embodied dataset," vol. 1, pp. 1–8, Apr. 2020, *arXiv:2004.13358*. [Online]. Available: http://arxiv.org/abs/2004.13358
[234] F. Sadeghi, "DIViS: Domain invariant visual servoing for collision-free goal reaching," vol. 1, pp. 1–11, Feb. 2019, *arXiv:1902.05947*. [Online]. Available: http://arxiv.org/abs/1902.05947
[235] C. Schenck and D. Fox, "Visual closed-loop control for pouring liquids," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Singapore, May 2017, pp. 2629–2636, doi: 10.1109/ICRA.2017.7989307.
[236] A. Ghadirzadeh, A. Maki, D. Kragic, and M. Bjorkman, "Deep predictive policy training using reinforcement learning," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Vancouver, BC, Canada, Dec. 2017, pp. 2351–2358, doi: 10.1109/IROS.2017.8206046.
[237] M. Andrychowicz, F. Wolski, A. Ray, J. Schneider, R. Fong, P. Welinder, B. McGrew, J. Tobin, O. P. Abbeel, and W. Zaremba, "Hindsight experience replay," in *Proc. Adv. Neur. Inf. Proc. Syst. (NIPS)*, 2017, pp. 5049–5059.
[238] A. Nair, V. Pong, M. Dalal, S. Bahl, S. Lin, and S. Levine, "Visual reinforcement learning with imagined goals," in *Proc. Adv. Neur. Inf. Proc. Syst. (NIPS)*, 2018, pp. 5049–5059.
[239] K. Katyal, I.-J. Wang, and P. Burlina, "Leveraging deep reinforcement learning for reaching robotic tasks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Honolulu, HI, USA, Jul. 2017, pp. 490–491, doi: 10.1109/CVPRW.2017.71.
[240] D. Kent, S. Banerjee, and S. Chernova, "Learning sequential decision tasks for robot manipulation with abstract Markov decision processes and demonstration-guided exploration," in *Proc. IEEE-RAS 18th Int. Conf. Hum. Robots (Humanoids)*, Beijing, China, Jan. 2018, pp. 958–965, doi: 10.1109/HUMANOIDS.2018.8624949.
[241] N. Gopalan, M. L. Littman, J. Macglashan, S. Squire, S. Tellex, and J. Winder, "Planning with abstract Markov decision processes," in *Proc. 27th Int. Conf. Automated Planning Scheduling*, Pittsburgh, PA, USA, Jun. 2017, pp. 480–488.
[242] M. Gualtieri, A. ten Pas, and R. Platt, "Pick and place without geometric object models," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Brisbane, QLD, Australia, Sep. 2018, pp. 7433–7440, doi: 10.1109/ICRA.2018.8460553.
[243] F. de La Bourdonnaye, C. Teuliere, T. Chateau, and J. Triesch, "Learning of binocular fixations using anomaly detection with deep reinforcement learning," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Anchorage, AK, USA, Jun. 2017, pp. 760–767, doi: 10.1109/IJCNN.2017.7965928.
[244] F. de La Bourdonnaye, C. Teulière, J. Triesch, and T. Chateau, "Stage-wise learning of reaching using little prior knowledge," *Frontiers Robot. AI*, vol. 5, p. 110, Oct. 2018, doi: 10.3389/frobt.2018.00110.
[245] F. de la Bourdonnaye, C. Teulière, T. Chateau, and J. Triesch, "Within reach? Learning to touch objects without prior models," in *Proc. Joint IEEE 9th Int. Conf. Develop. Learn. Epigenetic Robot. (ICDL-EpiRob)*, Oslo, Norway, Sep. 2019, pp. 93–98, doi: 10.1109/DEVLRN.2019.8850702.
[246] F. de La Bourdonnaye, C. Teulière, J. Triesch, and T. Chateau, "Learning to touch objects through stage-wise deep reinforcement learning," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Madrid, Spain, Jan. 2019, pp. 1–9, doi: 10.1109/IROS.2018.8593362.
[247] S. Wen, J. Chen, S. Wang, H. Zhang, and X. Hu, "Path planning of humanoid arm based on deep deterministic policy gradient," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Kuala Lumpur, Malaysia, Mar. 2019, pp. 1755–1760, doi: 10.1109/ROBIO.2018.8665248.
[248] B. Wu, K. Hao, X. Cai, X. Tang, and T. Wang, "Effective policy adjustment via meta-learning for complex manipulation tasks," in *Proc. Chin. Autom. Congr. (CAC)*, Xi'an, China, Jan. 2019, pp. 41–46, doi: 10.1109/CAC.2018.8623652.
[249] M. Breyer, F. Furrer, T. Novkovic, R. Siegwart, and J. Nieto, "Comparing task simplifications to learn closed-loop object picking using deep reinforcement learning," *IEEE Robot. Autom. Lett.*, vol. 4, no. 2, pp. 1549–1556, Apr. 2019, doi: 10.1109/LRA.2019.2896467.
[250] H. Xue, S. Wen, C. Yang, and H. Liu, "Model based reinforcement learning for robot grasping trajectory generation," in *Proc. IEEE 4th Int. Conf. Adv. Robot. Mechatronics (ICARM)*, Toyonaka, Japan, Jul. 2019, pp. 720–726, doi: 10.1109/ICARM.2019.8834254.
[251] J. P. King, D. Bauer, C. Schlagenhauf, K.-H. Chang, D. Moro, N. Pollard, and S. Coros, "Design. Fabrication, and evaluation of tendon-driven multi-fingered foam hands," in *Proc. IEEE-RAS 18th Int. Conf. Human Robots (Humanoids)*, Beijing, China, Jan. 2019, pp. 540–545, doi: 10.1109/HUMANOIDS.2018.8624997.
[252] C. Schlagenhauf, D. Bauer, K. Chang, J. P. King, D. Moro, S. Coros, and N. Pollard, "Control of tendon-driven soft foam robot hands," in *Proc. IEEE-RAS 18th Int. Conf. Hum. Robots (Humanoids)*, Beijing, China, Jan. 2018, pp. 1–7, doi: 10.1109/HUMANOIDS.2018.8624937.
[253] Y. She, S. Wang, S. Dong, N. Sunil, A. Rodriguez, and E. Adelson, "Cable manipulation with a tactile-reactive gripper," presented at the Robot., Sci. Syst. (RSS), Virtual Conf., Jul. 2020. [Online]. Available: https://roboticsconference.org/program/papers/29/ and http://arxiv.org/abs/1910.02860
[254] A. Sehgal, H. M. La, S. J. Louis, and H. Nguyen, "Deep reinforcement learning using genetic algorithm for parameter optimization," in *Proc. 3rd IEEE Int. Conf. Robot. Comput. (IRC)*, Naples, Italy, Mar. 2019, pp. 596–601, doi: 10.1109/IRC.2019.00121.
[255] C. Finn, X. Y. Tan, Y. Duan, T. Darrell, S. Levine, and P. Abbeel, "Deep spatial autoencoders for visuomotor learning," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, Stockholm, Sweden, Jun. 2016, pp. 512–519, doi: 10.1109/ICRA.2016.7487173.
[256] H. Liang, X. Lou, and C. Choi, "Knowledge induced deep Q-network for a slide-to-wall object grasping," 2019, pp. 1–7, *arXiv:1910.03781*. [Online]. Available: http://arxiv.org/abs/1910.03781
[257] T. Haarnoja, V. Pong, A. Zhou, M. Dalal, P. Abbeel, and S. Levine, "Composable deep reinforcement learning for robotic manipulation," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Brisbane, QLD, Australia, Sep. 2018, pp. 6244–6251, doi: 10.1109/ICRA.2018.8460756.
[258] M. Wu, Y. Gao, A. Jung, Q. Zhang, and S. Du, "The actor-dueling-critic method for reinforcement learning," *Sensors*, vol. 19, no. 7, p. 1547, Mar. 2019, doi: 10.3390/s19071547.
[259] I. Akinola, J. Varley, and D. Kalashnikov, "Learning precise 3D manipulation from multiple uncalibrated cameras," presented at the Int. Conf. Robot. Auto. (ICRA), Paris, France, Jul. 2020. [Online]. Available: http://arxiv.org/abs/2002.09107
[260] S. Gu, E. Holly, T. Lillicrap, and S. Levine, "Deep reinforcement learning for robotic manipulation with asynchronous off-policy updates," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Singapore, May 2017, pp. 3389–3396, doi: 10.1109/ICRA.2017.7989385.
[261] Y. Tsurumine, Y. Cui, E. Uchibe, and T. Matsubara, "Deep reinforcement learning with smooth policy update: Application to robotic cloth manipulation," *Robot. Auto. Syst.*, vol. 112, pp. 72–83, Feb. 2019, doi: 10.1016/j.robot.2018.11.004.
[262] T. Kobayashi, T. Aotani, J. R. Guadarrama-Olvera, E. Dean-Leon, and G. Cheng, "Reward-punishment actor-critic algorithm applying to robotic non-grasping manipulation," in *Proc. Joint IEEE 9th Int. Conf. Develop. Learn. Epigenetic Robot. (ICDL-EpiRob)*, Oslo, Norway, Sep. 2019, pp. 37–42, doi: 10.1109/DEVLRN.2019.8850699.

[263] L. Lv, S. Zhang, D. Ding, and Y. Wang, "Path planning via an improved DQN-based learning policy," *IEEE Access*, vol. 7, pp. 67319–67330, May 2019, doi: 10.1109/ACCESS.2019.2918703.

[264] C. C. Beltran-Hernandez, D. Petit, I. G. Ramirez-Alpizar, and K. Harada, "Learning to grasp with primitive shaped object policies," in *Proc. IEEE/SICE Int. Symp. Syst. Integr. (SII)*, Paris, France, Apr. 2019, pp. 468–473, doi: 10.1109/SII.2019.8700399.

[265] W. Montgomery and S. Levine, "Guided policy search via approximate mirror descent," in *Proc. Adv. Neur. Inf. Proc. Syst. (NIPS)*, 2016, pp. 4015–4023.

[266] F. Xiong, B. Sun, X. Yang, H. Qiao, K. Huang, A. Hussain, and Z. Liu, "Guided policy search for sequential multitask learning," *IEEE Trans. Syst. Man, Cybern., Syst.*, vol. 49, no. 1, pp. 216–226, Jan. 2019, doi: 10.1109/TSMC.2018.2800040.

[267] S. Demura, K. Sano, W. Nakajima, K. Nagahama, K. Takeshita, and K. Yamazaki, "Picking up one of the folded and stacked towels by a single arm robot," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Kuala Lumpur, Malaysia, Mar. 2019, pp. 1551–1556, doi: 10.1109/ROBIO.2018.8665040.

[268] M. P. Deisenroth, D. Fox, and C. E. Rasmussen, "Gaussian processes for data-efficient learning in robotics and control," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 2, pp. 408–423, Feb. 2015, doi: 10.1109/TPAMI.2013.218.

[269] H. Han, G. Paul, and T. Matsubara, "Model-based reinforcement learning approach for deformable linear object manipulation," in *Proc. 13th IEEE Conf. Automat. Sci. Eng. (CASE)*, Xi'an, China, Jan. 2017, pp. 750–755, doi: 10.1109/COASE.2017.8256194.

[270] R. B. Hellman, C. Tekin, M. van der Schaar, and V. J. Santos, "Functional contour-following via haptic perception and reinforcement learning," *IEEE Trans. Haptics*, vol. 11, no. 1, pp. 61–72, Mar. 2018, doi: 10.1109/TOH.2017.2753233.

[271] Q. Lei, J. Meijer, and M. Wisse, "A survey of unknown object grasping and our fast grasping algorithm-C shape grasping," in *Proc. 3rd Int. Conf. Control, Automat. Robot. (ICCAR)*, Nagoya, Japan, Jun. 2017, pp. 150–157, doi: 10.1109/ICCAR.2017.7942677.

[272] M. Alonso, A. Izaguirre, and M. Graña, "Current research trends in robot grasping and bin picking," in *Proc. Int. Joint Conf. SOCO-CISIS-ICEUTE SOCO-CISIS-ICEUTE*, vol. 771, M. Graña, Ed. San Sebastián, Spain: Springer, 2018, pp. 367–376, doi: 10.1007/978-3-319-94120-2_35.

[273] L. Pinto, J. Davidson, and A. Gupta, "Supervision via competition: Robot adversaries for learning tasks," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Singapore, Jul. 2017, pp. 1601–1608, doi: 10.1109/ICRA.2017.7989190.

[274] K. Fang, Y. Zhu, A. Garg, A. Kurenkov, V. Mehta, L. Fei-Fei, and S. Savarese, "Learning task-oriented grasping for tool manipulation from simulated self-supervision," *Int. J. Robot. Res.*, vol. 39, nos. 2–3, pp. 202–216, Aug. 2019, doi: 10.1177/0278364919872545.

[275] P. R. Florence, L. Manuelli, and R. Tedrake, "Dense object nets: Learning dense visual object descriptors by and for robotic manipulation," pp. 1–12, 2018, *arXiv:1806.08756*. [Online]. Available: https://arxiv.org/abs/1806.08756

[276] P. Agrawal, A. Nair, P. Abbeel, J. Malik, and S. Levine, "Learning to poke by poking: Experiential learning of intuitive physics," in *Proc. Adv. Neur. Inf. Proc. Syst. (NIPS)*, 2016, pp. 5074–5082.

[277] F. Ebert, C. Finn, S. Dasari, A. Xie, A. Lee, and S. Levine, "Visual foresight: Model-based deep reinforcement learning for vision-based robotic control," pp. 1–14, 2018, *arXiv:1812.00568v1*. [Online]. Available: https://arxiv.org/abs/1812.00568

[278] M. Zhang, S. Vikram, L. Smith, P. Abbeel, M. J. Johnson, and S. Levine, "Solar: Deep structured representations for model-based reinforcement learning," in *Proc. 36th Int. Conf. Mach. Learn. (PMLR)*, Jun. 2019, pp. 12861–12874.

[279] R. Chen and X. Dai, "Robotic grasp control policy with target predetection based on deep Q-learning," in *Proc. 3rd Int. Conf. Robot. Autom. Eng. (ICRAE)*, Guangzhou, China, Dec. 2018, pp. 29–33, doi: 10.1109/ICRAE.2018.8586758.

[280] A. Zeng, S. Song, J. Lee, A. Rodriguez, and T. Funkhouser, "TossingBot: Learning to throw arbitrary objects with residual physics," *IEEE Trans. Robot.*, vol. 36, no. 4, pp. 1307–1319, Aug. 2020, doi: 10.1109/TRO.2020.2988642.

[281] Y. Zaky, G. Paruthi, B. Tripp, and J. Bergstra, "Active perception and representation for robotic manipulation," pp. 1–8, 2020, *arXiv:2003.06734*. [Online]. Available: https://arxiv.org/abs/2003.06734

[282] D. Quillen, E. Jang, O. Nachum, C. Finn, J. Ibarz, and S. Levine, "Deep reinforcement learning for vision-based robotic grasping: A simulated comparative evaluation of off-policy methods," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, Brisbane, QLD, Australia, Sep. 2018, pp. 6284–6291, doi: 10.1109/ICRA.2018.8461039.

[283] W. Hu, C. Yang, K. Yuan, and Z. Li, "Reaching, grasping and re-grasping: Learning multimode grasping skills," pp. 1–8, 2020, *arXiv:2002.04498*. [Online]. Available: https://arxiv.org/abs/2002.04498

[284] Z. Xu, J. Wu, A. Zeng, J. Tenenbaum, and S. Song, "DensePhysNet: Learning dense physical object representations via multi-step dynamic interactions," pp. 1–10, 2019, *arXiv:1906.03853*. [Online]. Available: https://arxiv.org/abs/1906.03853

[285] H. Merzic, M. Bogdanovic, D. Kappler, L. Righetti, and J. Bohg, "Leveraging contact forces for learning to grasp," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, Montreal, QC, Canada, Aug. 2019, pp. 3615–3621, doi: 10.1109/ICRA.2019.8793733.

[286] Q. Lu, M. Van der Merwe, and T. Hermans, "Multi-fingered active grasp learning," pp. 1–9, 2020*arXiv:2006.05264*. [Online]. Available: https://arxiv.org/abs/2006.05264

[287] J. Ichnowski, M. Danielczuk, J. Xu, V. Satish, and K. Goldberg, "GOMP: Grasp-optimized motion planning for bin picking," pp. 1–8, 2020, *arXiv:2003.02401*. [Online]. Available: https://arxiv.org/abs/2003.02401

[288] J. Lundell, F. Verdoja, and V. Kyrki, "Beyond top-grasps through scene completion," pp. 1–7, 2019, *arXiv:1909.12908*. [Online]. Available: https://arxiv.org/abs/1909.12908

[289] R. Bemelmans, G. J. Gelderblom, P. Jonker, and L. de Witte, "Socially assistive robots in elderly care: A systematic review into effects and effectiveness," *J. Amer. Med. Directors Assoc.*, vol. 13, no. 2, pp. 114.e1–120.e1, 2012, doi: 10.1016/j.jamda.2010.10.002.

[290] R. Chitnis, D. Hadfield-Menell, A. Gupta, S. Srivastava, E. Groshev, C. Lin, and P. Abbeel, "Guided search for task and motion plans using learned heuristics," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Stockholm, Sweden, Jun. 2016, pp. 447–454, doi: 10.1109/ICRA.2016.7487165.

[291] B. Kim, L. P. Kaelbling, and T. Lozano-Perez, "Guiding the search in continuous state-action spaces by learning an action sampling distribution from off-target samples," pp. 1–8, 2017, *arXiv:1711.01391*. [Online]. Available: https://arxiv.org/abs/1711.01391

[292] Z. Wang, C. R. Garrett, L. P. Kaelbling, and T. Lozano-Perez, "Active model learning and diverse action sampling for task and motion planning," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Madrid, Spain, Jan. 2019, pp. 4107–4114, doi: 10.1109/IROS.2018.8594027.

[293] B. Kim, Z. Wang, L. P. Kaelbling, and T. Lozano-Pérez, "Learning to guide task and motion planning using score-space representation," *Int. J. Robot. Res.*, vol. 38, no. 7, pp. 793–812, Jun. 2019, doi: 10.1177/0278364919848837.

[294] R. Chitnis, L. P. Kaelbling, and T. Lozano-Perez, "Learning quickly to plan quickly using modular meta-learning," *Int. J. Robot. Res.*, vol. 38, no. 7, pp. 793–812, Jun. 2019, doi: 10.1177/0278364919848837.

[295] R. Silva, M. Faria, F. S. Melo, and M. Veloso, "Adaptive indirect control through communication in collaborative human-robot interaction," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Vancouver, BC, Canada, Dec. 2017, pp. 3617–3622, doi: 10.1109/IROS.2017.8206208.

[296] Z. Li, T. Zhao, F. Chen, Y. Hu, C. Su, and T. Fukuda, "Reinforcement learning of manipulation and grasping using dynamical movement primitives for a humanoidlike mobile manipulator," *IEEE/ASME Trans. Mechatronics*, vol. 23, no. 1, pp. 121–131, Feb. 2018, doi: 10.1109/TMECH.2017.2717461.

[297] H. Su, W. Qi, Y. Hu, J. Sandoval, L. Zhang, Y. Schmirander, G. Chen, A. Aliverti, A. Knoll, G. Ferrigno, and E. D. Momi, "Towards model-free tool dynamic identification and calibration using multilayer neural network," *Sensors*, vol. 19, no. 17, p. 3636, Aug. 2019, doi: 10.3390/s19173636.

[298] M. R. Maymó, A. Shafti, and A. A. Faisal, "Fast Orient: Lightweight computer vision for wrist control in assistive robotic grasping," in *Proc. 7th IEEE Int. Conf. Biomed. Robot. Biomechatronics (Biorob)*, Enschede, The Netherlands, Oct. 2018, pp. 207–212, doi: 10.1109/BIORobot.2018.8487622.

[299] H. Al-Fahaam, S. Davis, and S. Nefti-Meziani, "The design and mathematical modelling of novel extensor bending pneumatic artificial muscles (EBPAMs) for soft exoskeletons," *Robot. Auton. Syst.*, vol. 99, pp. 63–74, Jan. 2018, doi: 10.1016/j.robot.2017.10.010.

[300] J. Yi, X. Chen, and Z. Wang, "A three-dimensional-printed soft robotic glove with enhanced ergonomics and force capability," *IEEE Robot. Autom. Lett.*, vol. 3, no. 1, pp. 242–248, Jan. 2017, doi: 10.1109/LRA.2017.2737481.

[301] G. Grioli, A. Bicchi, F. Felici, M. G. Catalano, A. S. Ciullo, and A. Ajoudani, "Analytical and experimental analysis for position optimization of a grasp assistance supernumerary robotic hand," *IEEE Robot. Autom. Lett.*, vol. 3, no. 4, pp. 4305–4312, Oct. 2018, doi: 10.1109/LRA.2018.2864357.

[302] C. H. Corbato, M. Bharatheesha, J. van Egmond, J. Ju, and M. Wisse, "Integrating different levels of automation: Lessons from winning the amazon robotics challenge 2016," *IEEE Trans. Ind. Informat.*, vol. 14, no. 11, pp. 4916–4926, Nov. 2018, doi: 10.1109/TII.2018.2800744.

[303] R. Hodson, "How robots are grasping the art of gripping," *Nature*, vol. 557, no. 7704, pp. S23–S25, Aug. 2018, doi: 10.1038/d41586-018-05093-1.

[304] A. Kimmel, R. Shome, Z. Littlefield, and K. Bekris, "Fast, Anytime Motion Planning for Prehensile Manipulation in Clutter," in *Proc. IEEE-RAS 18th Int. Conf. Humanoid Robots (Humanoids)*, Beijing, China, Jan. 2018, pp. 1–9, doi: 10.1109/HUMANOIDS.2018.8624939.

[305] J. Shi and G. S. Koonjul, "Real-time grasping planning for robotic bin-picking and kitting applications," *IEEE Trans. Autom. Sci. Eng.*, vol. 14, no. 2, pp. 809–819, Apr. 2017, doi: 10.1109/TASE.2017.2671434.

[306] D. Angelov, Y. Hristov, M. Burke, and S. Ramamoorthy, "Composing diverse policies for temporally extended tasks," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 2658–2665, Apr. 2020, doi: 10.1109/LRA.2020.2972794.

[307] G. Thomas, M. Chien, A. Tamar, J. A. Ojea, and P. Abbeel, "Learning robotic assembly from CAD," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Brisbane, QLD, Australia, Sep. 2018, pp. 3524–3531, doi: 10.1109/ICRA.2018.8460696.

[308] Y. Litvak, A. Biess, and A. Bar-Hillel, "Learning pose estimation for high-precision robotic assembly using simulated depth images," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Montreal, QC, Canada, Aug. 2019, pp. 3521–3527, doi: 10.1109/ICRA.2019.8794226.

[309] A. Yamaguchi and C. G. Atkeson, "Stereo vision of liquid and particle flow for robot pouring," in *Proc. IEEE-RAS 16th Int. Conf. Humanoid Robots (Humanoids)*, Cancun, Mexico, Jan. 2017, pp. 1173–1180, doi: 10.1109/HUMANOIDS.2016.7803419.

[310] L. Rozo, P. Jiménez, and C. Torras, "Force-based robot learning of pouring skills using parametric hidden Markov models," in *Proc. 9th Int. Workshop Robot Motion Control*, Kuslin, Poland, Sep. 2013, pp. 227–232, doi: 10.1109/RoMoCo.2013.6614613.

[311] L. Kunze and M. Beetz, "Envisioning the qualitative effects of robot manipulation actions using simulation-based projections," *Artif. Intell.*, vol. 247, pp. 352–380, Jun. 2017, doi: 10.1016/j.artint.2014.12.004.

[312] A. Yamaguchi and C. G. Atkeson, "Differential dynamic programming with temporally decomposed dynamics," in *Proc. IEEE-RAS 15th Int. Conf. Humanoid Robots (Humanoids)*, Seoul, South Korea, Dec. 2015, pp. 696–703. 10.1109/HUMANOIDS.2015.7363430.

[313] C. Schenck, J. Tompson, D. Fox, and S. Levine, "Learning robotic manipulation of granular media," in *Proc. 1st Annu. Conf. Robot Learn. (CoRL)*, Mountain View, CA, USA, Nov. 2017, pp. 239–248.

[314] A. Yamaguchi and C. G. Atkeson, "Differential dynamic programming for graph-structured dynamical systems: Generalization of pouring behavior with different skills," in *Proc. IEEE-RAS 16th Int. Conf. Humanoid Robots (Humanoids)*, Cancun, Mexico, Jan. 2017, pp. 1029–1036, doi: 10.1109/HUMANOIDS.2016.7803398.

[315] C. Schenck and D. Fox, "Reasoning about liquids via closed-loop simulation," pp. 1–10, 2017. *arXiv:1703.01656v2*. [Online]. Available: https://arxiv.org/abs/1703.01656

[316] A. Yamaguchi and C. G. Atkeson, "Neural networks and differential dynamic programming for reinforcement learning problems," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Stockholm, Sweden, Jun. 2016, pp. 5434–5441, doi: 10.1109/ICRA.2016.7487755.

[317] F. Lévesque, B. Sauvet, P. Cardou, and C. Gosselin, "A model-based scooping grasp for the autonomous picking of unknown objects with a two-fingered gripper," *Robot. Auton. Syst.*, vol. 106, pp. 14–25, Aug. 2018, doi: 10.1016/j.robot.2018.04.003.

[318] V. Babin, D. St-onge, and C. Gosselin, "Stable and repeatable grasping of fl at objects on hard surfaces using passive and epicyclic mechanisms," *Robot. Comput. Integr. Manuf.*, vol. 55, pp. 1–10, Feb. 2019, doi: 10.1016/j.rcim.2018.06.002.

[319] V. Babin and C. Gosselin, "Picking, grasping, or scooping small objects lying on flat surfaces: A design approach," *Int. J. Robot. Res.*, vol. 37, no. 12, pp. 1484–1499, Oct. 2018, doi: 10.1177/0278364918802346.

[320] H. Bharadhwaj, S. Yamaguchi, and S. Maeda, "MANGA: Method agnostic neural-policy generalization and adaptation," pp. 1–8, 2019. *arXiv:1911.08444v1*. [Online]. Available: https://arxiv.org/abs/1911.08444

[321] A. X. Lee, R. Zhang, F. Ebert, P. Abbeel, C. Finn, and S. Levine, "Stochastic adversarial video prediction," pp. 1–26, 2018. *arXiv:1804.01523v1*. [Online]. Available: https://arxiv.org/abs/1804.01523

[322] A. Nair, S. Bahl, A. Khazatsky, V. Pong, G. Berseth, and S. Levine, "Contextual imagined goals for self-supervised robotic learning," in *Proc. 3rd Conf. Robot Learn. (CoRL)*, Osaka, Japan, Oct./Nov. 2019, pp. 530–539.

[323] R. Rahmatizadeh, P. Abolghasemi, L. Boloni, and S. Levine, "Vision-based multi-task manipulation for inexpensive robots using end-to-end learning from demonstration," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Brisbane, QLD, Australia, Sep. 2018, pp. 3758–3765, doi: 10.1109/ICRA.2018.8461076.

MARWAN QAID MOHAMMED received the B.Eng. (Hons.) and M.Eng.Sc. degrees in electrical engineering (mechatronics) from Universiti Teknikal Malaysia Melaka (UTeM), in 2016 and 2018, respectively. He is currently a Graduate Research Assistance (GRA) with Multimedia University (MMU). His main research interests include trajectory generation, motion recognition, robotics, and reinforcement learning algorithms.

KWEK LEE CHUNG (Member, IEEE) received the B.Tech. degree in industrial technology from the University Science of Malaysia and the Master of Engineering Science and Ph.D. degrees in engineering from Multimedia University, Malaysia. She is currently a Lecturer with the Faculty of Engineering and Technology, Multimedia University. Her research interests include robotic controls and artificial intelligent methods.

CHUA SHING CHYI (Senior Member, IEEE) received the B.Eng. (Hons.), M.Eng.Sc., and Ph.D. degrees in electronics from Multimedia University (MMU). He is currently a Senior Lecturer with the Faculty of Engineering and Technology (FET), MMU. His research interests include robotics, fuzzy logic, and digital image processing.