


Article

Increasing the Energy-Efficiency in Vacuum-Based Package Handling Using Deep Q-Learning

Felix Gabriel *, Johannes Bergers, Franziska Aschersleben  and Klaus Dröder

Institute of Machine Tools and Production Technology, Technische Universität Braunschweig, Langer Kamp 19b, 38106 Braunschweig, Germany; j.bergers@tu-braunschweig.de (J.B.); f.aschersleben@tu-braunschweig.de (F.A.); k.droeder@tu-braunschweig.de (K.D.)

* Correspondence: f.gabriel@tu-braunschweig.de

Abstract: Billions of packages are automatically handled in warehouses every year. The gripping systems are, however, most often oversized in order to cover a large range of different carton types, package masses, and robot motions. In addition, a targeted optimization of the process parameters with the aim of reducing the oversizing requires prior knowledge, personnel resources, and experience. This paper investigates whether the energy-efficiency in vacuum-based package handling can be increased without the need for prior knowledge of optimal process parameters. The core method comprises the variation of the input pressure for the vacuum ejector, compliant to the robot trajectory and the resulting inertial forces at the gripper-object-interface. The control mechanism is trained by applying reinforcement learning with a deep Q-agent. In the proposed use case, the energy-efficiency can be increased by up to 70% within a few hours of learning. It is also demonstrated that the generalization capability with regard to multiple different robot trajectories is achievable. In the future, the industrial applicability can be enhanced by deployment of the deep Q-agent in a decentral system, to collect data from different pick and place processes and enable a generalizable and scalable solution for energy-efficient vacuum-based handling in warehouse automation.



Citation: Gabriel, F.; Bergers, J.; Aschersleben, F.; Dröder, K. Increasing the Energy-Efficiency in Vacuum-Based Package Handling Using Deep Q-Learning. *Energies* **2021**, *14*, 3185. <https://doi.org/10.3390/en14113185>

Academic Editor: Satoru Okamoto

Received: 26 April 2021

Accepted: 27 May 2021

Published: 29 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: vacuum-based handling; energy-efficiency; deep Q-learning; automation

1. Introduction

Vacuum-based handling is used in a large variety of applications, especially when a high flexibility is required due to diverse objects that must be grasped, e.g., in packaging and warehouse logistics. These fields of application are constantly gaining relevance, as global retail e-commerce sales amounted to 4.3 trillion US dollars in 2020 and revenues are estimated to grow to 6.4 trillion dollars until 2024 [1]. Current vacuum-based gripping systems for package handling are mostly realized by means of compressed air-supplied vacuum ejectors and therefore exhibit a highly dynamic and wear-free operation. However, using vacuum ejectors causes enormous energy losses [2,3], since only a few percent of the initially provided electrical energy can be utilized for the handling process (Figure 1). Hence, it is crucial to design the gripping system and the corresponding process parameters in compliance with the application-specific requirements such as the expected robot trajectories and the properties of the objects to be handled. Since handling tasks are not value-added [4], in industrial practice, it is crucial to setup the system and process fast. In particular, in case of a large spectrum of objects to be handled, it is usually not economically feasible to adjust the process parameters for each specific trajectory and object. Hence, in practice, a universally applicable standard system is set up which is oversized for most of the expected objects, but will eventually provide a robust handling process.

These standard systems are normally dimensioned in accordance with basic calculation schemes, under consideration of the most prevalent load case [5]. Based on the maximum expected load, aggregated by gravitational and inertial forces, the required

number and size of vacuum grippers are selected, as well as the necessary pressure difference and a sufficiently powerful ejector type. Due to manifold uncertainties such as environmental conditions (temperature, humidity, contamination), exact object properties (carton composition, mass) and gripper behavior (deformation, sealing capabilities), a certain safety margin is finally applied in order to oversize the calculated system and process parameters. Once these parameters are set, the possibility of online adjustments during the running process is limited to variation of the input pressure of compressed air, which the vacuum ejector is supplied with. However, it would require a disproportionately high effort to analyze every possible load scenario by hand and prior to getting the handling process running.

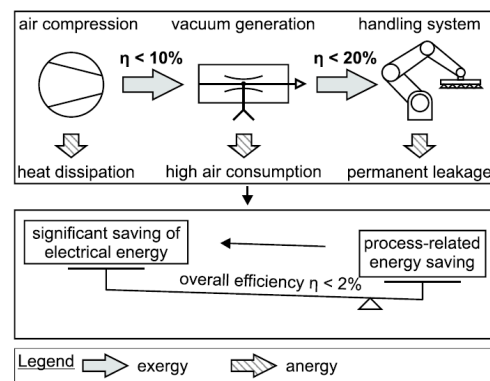


Figure 1. Energy conversion losses in compressed air-based vacuum generation [2].

Aiming at improving the energy-efficiency in vacuum-based handling, several papers can be found on development of improved or novel vacuum grippers. Most common approaches focus on the integration of shape memory alloy in order to actively control the adaptation capabilities of single suction cups [6–9]; other researchers develop biomimetic vacuum grippers [10], origami-inspired [11], or electrically actuated grippers [12,13]. Extensive work can be found on mathematical modeling of vacuum grippers with the objective of a more precise system and/or process dimensioning. Basis static modeling of vacuum grippers is conducted in [11,13–21], dynamic model approaches are presented in [2,22–24]. A few publications can be found that focus on finite element analysis and, based on that, design optimization of vacuum grippers [25,26]. Another field concentrates on finding the optimal gasp points of multiple vacuum grippers on the part surface [27–30]. With regard to fluid dynamics and acoustics, several publications on optimized vacuum ejectors and enhanced air-saving functionalities are present, as well [31–33]. The majority of the here-presented related work focuses on grasping air-impermeable objects and is therefore not directly applicable to package handling. In the context of carton or package handling, vacuum grippers are rather utilized as supportive elements in flexible gripping systems for depalletizing [34,35]. Methods specializing on improving the energy-efficiency for vacuum-based handling of air-permeable objects such as packages or textiles can hardly be found in literature. The application of machine learning methods in vacuum-based handling is fairly limited in literature. In [36], an adaptive vacuum monitoring and control system for a passive vacuum generation mechanism is realized by means of a deep Q-agent (DQA). Mahler et al. predict the robustness of a vacuum suction grasp via an analytic model and a deep Q-learning (DQL) algorithm [28]. With regard to bin picking, however, a huge body of research is present. Based on image data, the most feasible grasp pose is estimated in order to pick objects from a bin [37–41]. Clearly, the application of reinforcement learning (RL) in vacuum-based handling has already shown great potential for technological improvements. In summary, there are a number of methods for increasing energy efficiency in vacuum-based handling, but gripping air-permeable objects is a major challenge that can be improved using DQL.

Therefore, the aim of this paper is to investigate whether the energy-efficiency in vacuum-based package handling can be increased without the need for prior knowledge of optimal process parameters. The core idea is to realize the adaptive variation of the input pressure for the vacuum ejector, compliant to the robot trajectory and the resulting inertial forces at the gripper-object-interface, based on a DQL approach. With this approach, extensive prior knowledge about process-specific parameters that are advantageous for both a robust and an efficient handling process can be made obsolete. In several subsequent training episodes, a deep Q-agent trains to predict the impact of a certain pressure profile in combination with a specific robot trajectory (the object is not varied). This paper examines how fast and to what extent the DQA learns to improve the energy efficiency and how well it is able to generalize with regard to different trajectories. It is demonstrated that energy savings of 50% to almost 70% can be achieved within an hour of iterative experiments. It is also shown that a good generalization capability of the DQA with regard to multiple different robot trajectories can be achieved.

The paper is structured as follows. Section 2 discusses materials and methods, i.e., the utilized experimental setup as well as the underlying method and the implementation of the proposed DQA approach. The results of the conducted experimental case study are demonstrated in Section 3. Finally, in Section 4, the results are discussed with regard to industrial applicability and scalability and conclusions for further research are drawn.

2. Materials and Methods

In this section, the regarded use case is firstly presented and the corresponding experimental setup is introduced. Subsequently, the conceptual design and implementation of the deep Q-agent is elaborated in combination with the process control system.

2.1. Use Case Definition for Vacuum-Based Package Handling and Experimental Setup

An industrially relevant use case was initially defined in order to create a realistic set of process requirements in accordance to typical package handling processes in warehouse automation. In addition, the transferability of the obtained findings into industrial applications can thus be estimated. In the scope of the underlying research project BiVaS, an experimental robot-based setup is available at Open Hybrid LabFactory (OHLF) in Wolfsburg, Germany. This robot-based setup (Figure 2a), is supplied with centrally generated compressed air and mainly comprises a vacuum ejector, a proportional valve for variation of the input pressure, a gripping system, and a distance sensor for detection of object presence. The gripping system was supplied by J. Schmalz GmbH and consists of four identical vacuum grippers of type SPB1 60 ED-65. The process control is realized as soft PLC on a control PC (Beckhoff TwinCAT 3). Figure 2b shows the target positions of the realized use case. For the robot movement from start to end position, a maximum duration of 2.5 s was defined in accordance with typical industrial handling applications with about 10–15 picks per minute (information provided by J. Schmalz GmbH, Glatten, Germany).

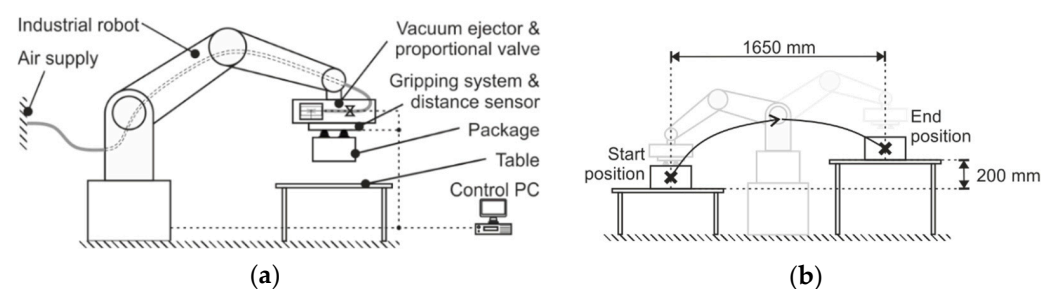


Figure 2. (a) Experimental robot-based setup at Open Hybrid LabFactory (OHLF) [42] in Wolfsburg, Germany; (b) target positions for the defined use case.

The straightforward approach to save energy in this specified use case is to reduce the input pressure via the proportional valve and therefore decrease the compressed air consumption. In case of setting the pressure too low, the package may fall off the gripping system. In order to evaluate whether the package has lost contact with the grippers, a distance sensor is integrated into the gripping system (Figure 3a). If a certain threshold value for the measured distance is exceeded, the package is considered fallen off. In order to enable fully automated tests, the package was fixed to the gripping system with belts (Figure 3b). This makes it possible for the robot to carry the package back to the start position and start over the handling process, regardless of the process success.

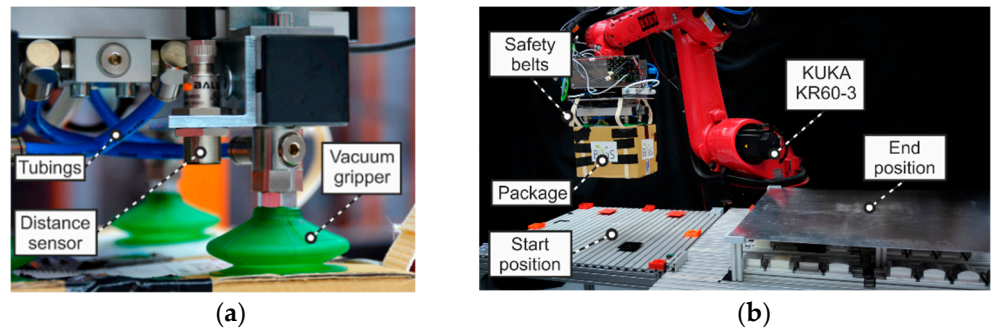


Figure 3. (a) An ultrasonic distance sensor detects if the gripping system loses the package; (b) photo of test setup at OHLEF.

Based on the process specification (Figure 2b), 18 different trajectories were created in order to generate multiple diverse acceleration profiles in X- and Z-component (in the scope of this work, the robot path is regarded two-dimensional). Specifically, six different robot paths were created by first defining a spline-based path with subsequent variation of the support points, and then applied with three different execution speeds each. Figure 4 shows the created paths in X- and Z-components for all 18 trajectories T_1 to T_{18} . The objective of the experimental case study is to evaluate as a first step if a trained DQA is able to increase the energy efficiency of this handling set-up. As a second step, the quality of the DQA is analyzed with respect to the ability to reduce the energy consumption for each separate trajectory and, in comparison, how well it is capable of generalizing and transferring the learned mechanism to unknown trajectories.

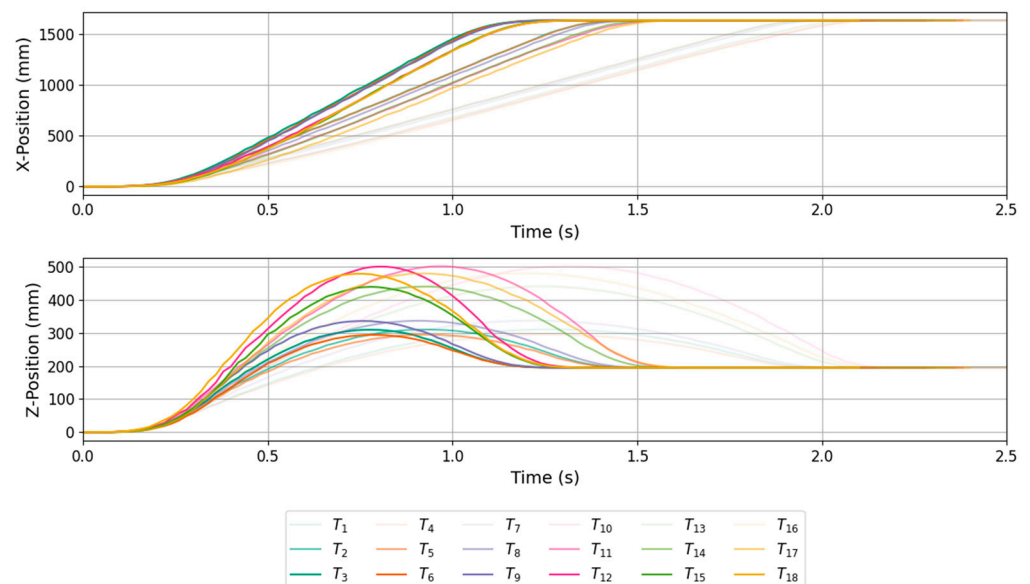


Figure 4. The designed trajectories result in diverse acceleration profiles.

2.2. Deep Q-Learning Implementation and Process Control Architecture

Reinforcement Learning (RL) is one of three main categories of Machine Learning besides Supervised Learning, which is primarily applied for classification and regression, and Unsupervised Learning, most often used for cluster recognition and data compression [43]. RL aims to train an agent in making decisions in order to reach a certain defined objective by maximizing the reward that depends on the outcome of each episode (and each step inside one episode). The agent interacts with the environment by performing actions based on observations of states. At each time step t , the system is in a state s_t . Originating from this state, the agent picks a certain action a_t from the space of possible actions A . This action leads to a new state s_{t+1} which is associated with a reward r_{t+1} . For a given number of episodes, the agent learns a strategy to decide on the optimal state-specific action, called policy. The core idea of RL is based on describing the regarded optimization problem as Markov-Decision-Process (MDP) problem, which assumes a finite number of states and actions, where each subsequent state depends only on the current state [44]. One established approach to solve the MDP problem is Q-learning which is a model-free method based on Temporal Difference Learning. Q-learning offers a high sample efficiency which is particularly advantageous for practical experiments [45]. The objective of Q-learning is to learn to decide which action will yield the highest reward, depending on the state. The optimal Q-value Q^* is hereby calculated by Equation (1):

$$Q^*(s_t, a_t) = R(s_t, a_t) + \gamma \max_{a \in A} Q(s_{t+1}, a_{t+1}). \quad (1)$$

R is the expected total reward for the current state-action pair (s_t, a_t) . In addition to R , the maximum expectable Q-value of all possible next state-action pairs is estimated and discounted over time by the discount factor γ . In each training episode, the Q-learning algorithm updates the approximation of the Q-function with Equation (2), where α is the learning rate.

$$Q^{\text{upd}}(s_t, a_t) = (1 - \alpha) Q(s_t, a_t) + \alpha \left[r(s_t, a_t) + \gamma \max_{a \in A} Q(s_{t+1}, a_{t+1}) \right] \quad (2)$$

Figure 5 visualizes the implementation of the DQA approach for the above specified use case. The input for the DQA, the state s_t , is defined by the acceleration profile. With the state as input, the deep Q-agent predicts the optimal Q-value and thus selects the best action which represents discrete pressure levels.

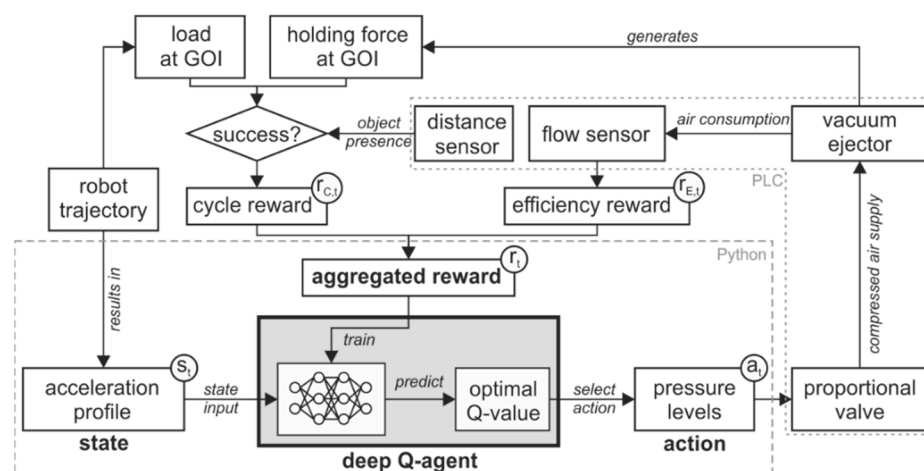


Figure 5. Implementation of the DQA for the proposed vacuum-based handling problem.

For the execution of the handling process, the process control system (PLC) sets the pressure levels to the proportional valve at pre-defined times. This timed pressure control

supplies the vacuum ejector with (ideally) the exact amount of pressurized air that is needed for a robust but efficient handling process, in order to encounter the load occurring at the gripper-object-interface (GOI) with an adequately high holding force. The efficiency reward is computed by the measurement of the consumed air flow and combined with the cycle reward (success or failure) to the aggregated reward r_t . This allows the agent to be trained with the newly generated dataset of state, action, and reward. The DQA (two fully connected hidden layers of 24 neurons) is implemented by means of the Keras framework and trained with a learning rate of 0.01.

To condensate the robot trajectory information into a state representation that is usable for the DQA, it is required to evaluate the respective trajectory with regard to the most relevant acceleration values. Figure 6 presents a method for feature extraction from acceleration profiles in pick and place processes. In general, typical pick and place trajectories mainly consist of a motion in Z which goes up initially, when an object is picked, and goes down in the end, when the object is placed. The resulting path roughly follows the form of a parabola (Figure 6 top, dashed line). The motion in X follows the shape of a sigmoid curve, initially accelerating to a constant velocity and decelerating in the end (Figure 6 top, dotted line). Since these motion patterns can be assumed to happen in most pick and place processes as shown qualitatively in Figure 6, it is possible to derive a generally applicable method for identifying the time windows where the input pressure should be varied. Regarding the resulting accelerations profiles of the X- and Z-motion, three points of pairs of interest (POI) can be defined (marked green in Figure 6, bottom), that are relevant for choosing the appropriate pressure level in accordance with common load cases (normal load, transversal load).

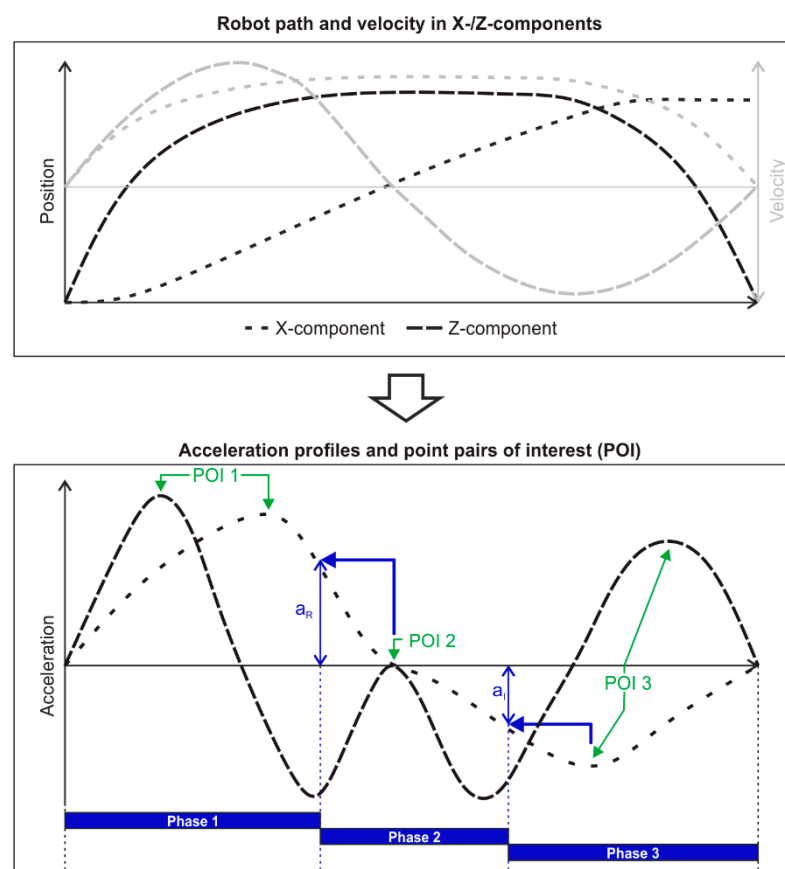


Figure 6. Method for feature extraction from acceleration profiles in pick and place processes.

This results in three process phases, where different pressure levels can be set. The first POI is associated with the initial acceleration of the robot, where a positive acceleration

happens in both the X- and Z-component. The second POI is located at the transition from positive to negative acceleration in X. In this area, the Z-acceleration becomes zero. Hence, the input pressure (and therefore the holding force) can potentially be reduced. To decide at what time the pressure may be reduced after the initial acceleration, it is first necessary to assess what residual acceleration a_R is permissible. This defines the duration of Phase 1 (blue left-headed arrow in Figure 6). The third POI covers the maximum deceleration of the X-motion and the deceleration of Z approaching the target position for deposition of the object. Accordingly, based on the value of the initial acceleration a_1 at that point when the input pressure should be increased again, the duration of Phase 2 can be determined. Finally, the duration of Phase 3 is set, as well, since it lasts until the end of the process. For the application of Q-learning to the above-defined problem, it is required to define finite state and action spaces. Hence, discrete acceleration features were derived from each of the designed robot trajectories. For each of the defined phases, the maximum acceleration values for X, -Z, and +Z were determined and rounded to one decimal place. One set of these three acceleration values composes one state $s_t = [a_{x,\max}, a_{+z,\max}, a_{-z,\max}]$. The maximum obtained acceleration values accounted for $a_{x,\max} = 10.0 \text{ m/s}^2$, $a_{+z,\max} = 6.2 \text{ m/s}^2$ and $a_{-z,\max} = 5.2 \text{ m/s}^2$, which therefore results in an overall space of $100 \times 62 \times 52 = 322,400$ three-dimensional possible states in the specified use case scenario (however, not all of these states are physically possible).

In each of the three phases, based on the specific state, the agent decides for an input pressure between 0.0 and 6.0 bar, in steps of 0.25 bar. Thus, the action space contains 25 possible actions. For each episode (one complete handling process from start to end position), based on the phase-specific states, three actions are offline composed to one set and then executed. In the training process, the resulting efficiency rewards can be directly associated with each phase. The efficiency reward is calculated by dividing the compressed air saving, hence the difference between the reference air consumption c_r and the measured compressed air consumption c_m , by the reference air consumption c_r , which was previously determined for each trajectory and a permanent pressure level of 6.0 bar. This reference case originates from industrial practice, since in most cases, a compressed air network is already present in the factory and supplies the air at 6–8 bar. For the first two phases, the reward is calculated as follows:

$$r_E = \frac{c_r - c_m}{c_r} \quad (3)$$

For the last phase, it is additionally considered whether the handling process has been completed properly. The reward is calculated by

$$r = r_E + r_C, \quad (4)$$

where r_C is the cycle reward which is set to 1 in case of a success and to -10 in case of failure. This enables to not consider the previous energy savings isolated but also their contribution to the process failure. If the package is lost in Phase 1 or Phase 2, the reward is directly set to zero for the respective phase. The state representations that were determined experimentally by the created trajectories and the proposed method for extraction of the acceleration features introduced in Figure 6, as well as the corresponding reference air consumption values for each trajectory (given in liters), are summarized in Table 1.

Table 1. State representations and corresponding reference air consumption for each trajectory.

Trajectory No.	States [$a_{x,max}$, $a_{+z,max}$, $a_{-z,max}$] (m/s^2)	Reference Air Consumption (L)
1	[5.3, 2.5, 0.0], [0.3, 0.0, 0.6], [5.8, 1.6, 0.0]	3.262
2	[7.7, 3.6, 0.0], [0.6, 0.0, 1.5], [8.0, 2.1, 0.9]	2.507
3	[9.2, 4.3, 0.0], [0.8, 0.0, 2.6], [9.9, 2.3, 0.0]	2.114
4	[5.3, 2.4, 0.0], [0.3, 0.0, 0.6], [5.8, 1.4, 0.0]	3.262
5	[7.4, 3.7, 0.0], [0.6, 0.0, 1.4], [8.0, 1.9, 0.0]	2.507
6	[9.1, 4.0, 0.0], [0.6, 0.0, 2.3], [10.0, 2.0, 0.0]	2.144
7	[5.0, 2.9, 0.0], [0.4, 0.0, 0.7], [5.6, 1.8, 0.0]	3.292
8	[7.0, 4.1, 0.0], [0.7, 0.0, 1.6], [8.0, 2.2, 0.0]	2.537
9	[8.9, 5.0, 0.0], [0.4, 0.0, 2.8], [9.9, 2.6, 0.0]	2.144
10	[4.4, 3.5, 0.0], [0.4, 0.0, 1.3], [4.6, 3.5, 0.0]	3.533
11	[6.0, 5.0, 0.0], [1.4, 0.0, 2.9], [6.6, 4.6, 0.0]	2.688
12	[7.4, 5.8, 0.0], [1.4, 0.0, 5.2], [8.4, 5.5, 0.0]	2.265
13	[4.3, 3.7, 0.0], [0.5, 0.0, 1.0], [5.2, 2.6, 0.0]	3.443
14	[5.9, 4.8, 0.0], [0.9, 0.0, 0.9], [7.1, 3.5, 0.0]	2.597
15	[7.7, 5.8, 0.0], [0.6, 0.0, 3.9], [9.7, 4.4, 0.0]	2.235
16	[3.8, 3.9, 0.0], [1.2, 0.0, 1.2], [5.1, 2.8, 0.0]	3.473
17	[5.4, 5.4, 0.0], [1.2, 0.0, 2.5], [7.2, 3.6, 0.0]	2.688
18	[6.6, 6.3, 0.0], [2.4, 0.0, 4.5], [9.4, 4.4, 0.0]	2.265

2.3. Design and Conduction of Experiments

Two experimental studies were designed and conducted in order to evaluate the capability of the DQA to reduce the energy consumption of the introduced use case. The first objective is to train an agent isolated for each trajectory. Hence, all 18 trajectories were implemented as subprograms in the robot control system and then executed in a main program accessed via indices. For the first experiment, 700 repetitions were planned for each trajectory. The indices were shuffled for randomization. In total, 12,600 repetitions were conducted in a duration of about 34 h. The E-Greedy algorithm with an exploration decay rate of $\epsilon = 0.999$ was applied to initially ensure a sufficiently strong exploration behavior. In comparison to the Greedy algorithm, E-Greedy balances exploration and exploitation by means of the exploration decay rate. This means that at a probability of ϵ , a set of random actions is selected which is different from the action set that was selected by the DQA. Over time, this randomness decays since in each n -th episode, the probability for a random action set accounts for ϵ^n (with $\epsilon_{min} = 0.01$).

The second experiment aims to evaluate the generalization capabilities of the DQA. The underlying idea is to train the DQA with data from 15 out of 18 trajectories (training data) in a total of 10,000 repetitions, again with $\epsilon = 0.999$. Subsequently, the so-trained neural network is re-used as pre-trained DQA and applied to the remaining three trajectories (test data), that are so far unknown to the DQA, and trained with the acquired data. For 2000 additional repetitions, the exploration decay rate is set to $\epsilon = 0$. For this second experiment, a 6-fold cross validation was planned in order to compare the generalization results for different combinations of training and test data sets. Since the trajectories have been generated by varying six different paths by three different values for the program override, three categories of trajectories are available: slow, medium, and fast. For selecting indices for the cross validation, it was ensured that the test indices contain exactly one of each category. Table 2 assigns the trajectories to the respective categories.

Table 2. Categories of created trajectories for the design of cross validation experiments.

Category	Trajectory Indices
slow	1,4,7,10,13,16
medium	2,5,8,11,14,17
fast	3,6,9,12,15,18

Accordingly, the training and test indices were selected for the 6-fold cross validation. For each of the six folds, lists of 10,000 slots were filled by random selection out of the respective training indices (see Table 3), and analogous for the test indices. For each fold, the experiments took about 33 h (12,000 repetitions at ~360 reps/h) in total.

Table 3. Training and test indices for 6-fold cross validation of the second experiment.

Fold No.	Training indices	Test Indices
1	1,2,3,4,5,6,7,9,11,12,13,14,15,16,17	8,10,18
2	1,2,3,4,5,7,8,9,10,11,12,14,15,16,18	6,13,17
3	1,3,4,5,6,8,9,10,11,13,14,15,16,17,18	2,7,12
4	2,4,5,6,7,8,9,10,11,12,13,15,16,17,18	1,3,14
5	1,2,3,4,6,7,8,9,10,11,12,13,14,17,18	5,15,16
6	1,2,3,5,6,7,8,10,12,13,14,15,16,17,18	4,9,11

3. Results

The results of the first experiments are shown in Figure 7. For better visibility, the running mean was computed with a range of 50 data points. In case of the slow and medium trajectories (columns 1 and 2), the reward quickly approaches the value of 3 within 400 episodes (~1 h). For the fast trajectories (column 3), the results are not that clear; the reward even decreases in two cases.

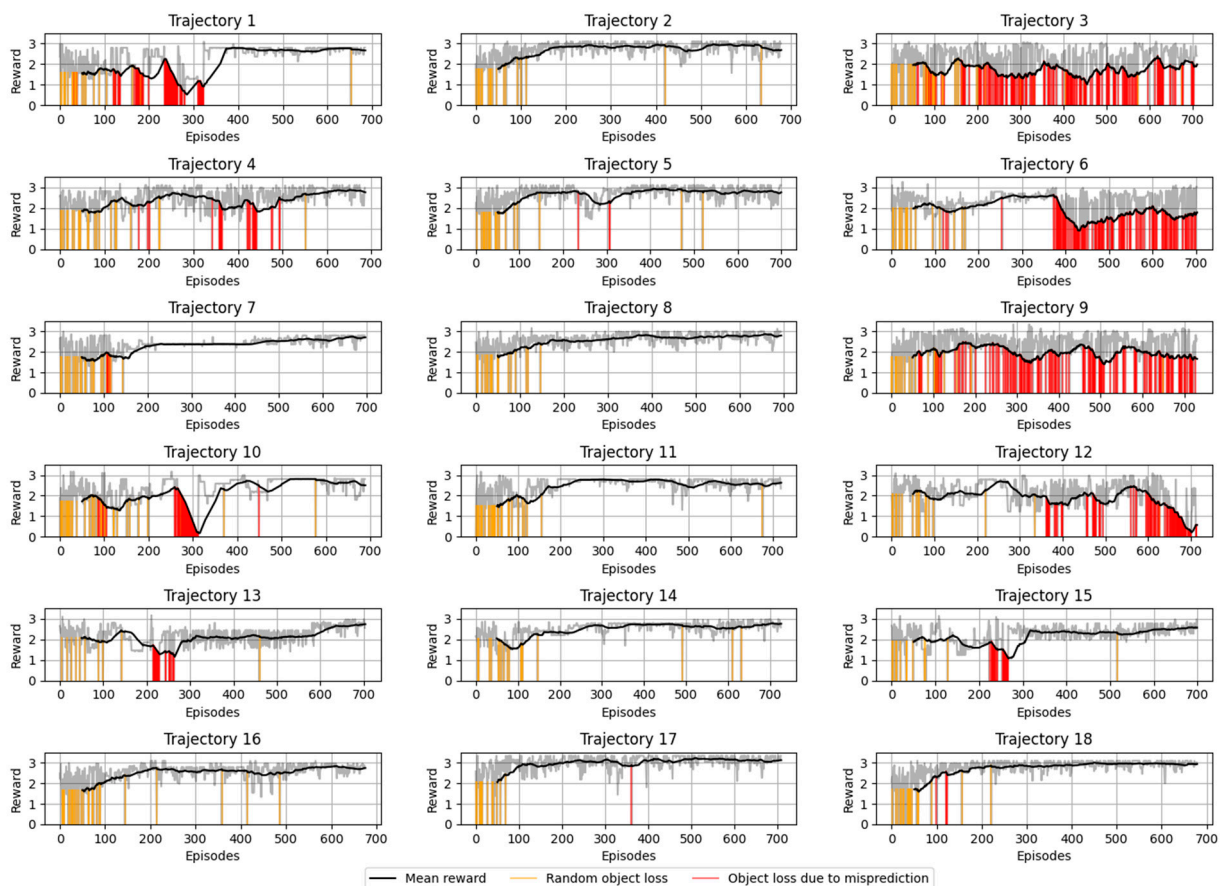


Figure 7. Experiment 1: reward over episodes for all 18 trajectories.

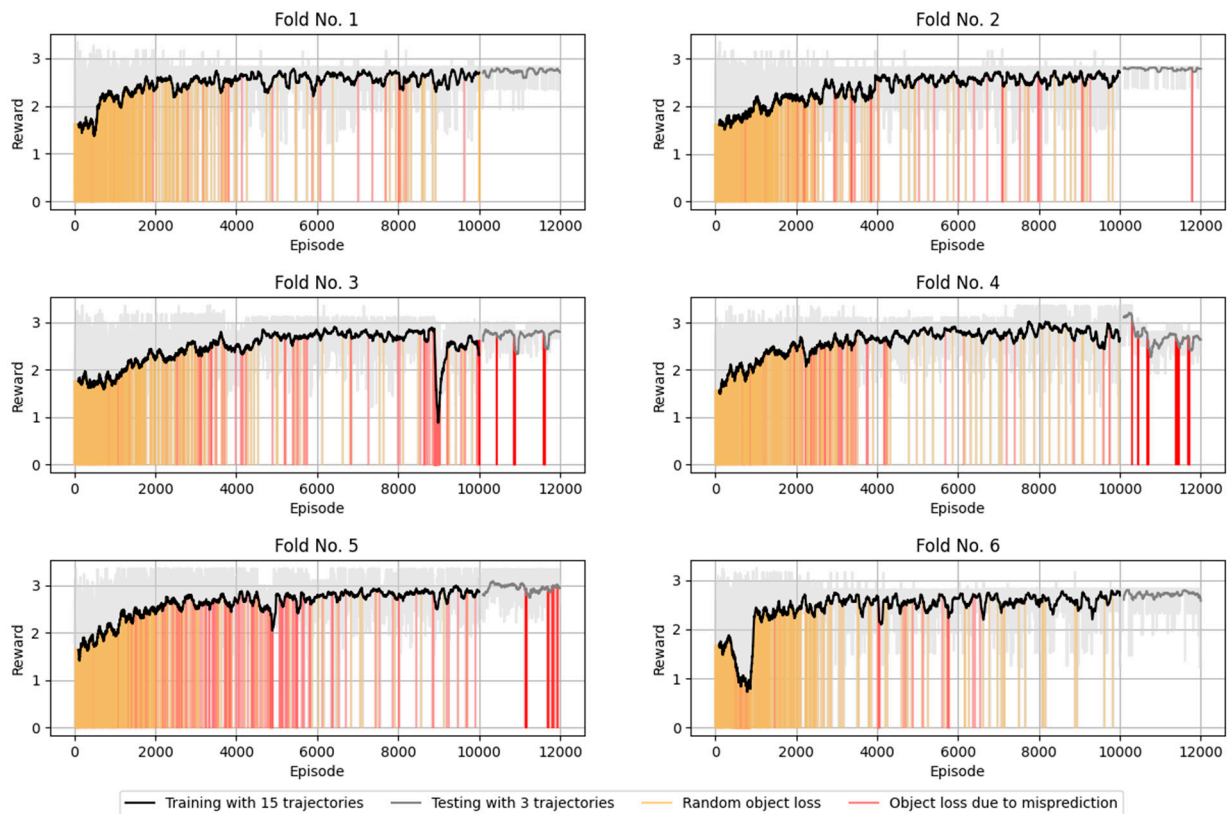
According to the introduced reward function, average energy savings of 50% up to about 70% can be achieved fast (compare Table 4).

Table 4. Relation between reward and compressed air saving for exemplary rewards.

Reward (Efficiency Reward + Cycle Reward)	Compressed Air Saving
1.75 (0.20 + 0.25 + 0.30 + 1)	25%
2.5 (0.6 + 0.4 + 0.5 + 1)	50%
3.25 (0.75 + 0.75 + 0.75 + 1)	75%

In general, two different cases of object loss happen. Firstly, a random object loss occurs when the exploration rate is sufficiently high for the DQA to select an action different from the predicted optimal action. For example, setting the pressure to zero in at least one of the three process phases will eventually lead to an object loss. Secondly, a misprediction can also lead to an object loss. It is noticeable that especially in the case of the medium trajectories, almost no object losses due to mispredictions occur. Conversely, for the trajectories 3, 6, 9, and 12, a high amount of mispredictions detracts the agent from converging at all. In the example of trajectory 6, early stopping at approx. episode 380 would have led to a reward of about 2.5. The following training episodes seem to force the agent into overfitting.

The results of the second experiment with 6-fold cross validation are depicted in Figure 8. In general, the reward converges towards a value of 3 during training as the amount of object losses due to mispredictions decreases over time. An exception is the experiment associated with fold no. 3, where a significant accumulation of such events occurs around episode 9,000. With the pre-trained neural network, in all folds except no. 4, rewards are quickly reached that are either in the range of the final training reward or even higher (e.g., in fold no. 2). Fold no. 4 is another exception in this case, where several mispredictions lead to a more conservative behavior of the DQA.

**Figure 8.** Results of the 6-fold cross validation experiments for evaluation of the generalization capability of the DQA.

4. Discussion

For both experiments conducted, a simple DQA was set up and trained with the acquired data. Without extensive tuning of hyper parameters, particularly for slow and medium trajectories, energy savings of 50% (reward of 2.5) to almost 70% (reward of 3.1) were achieved within an hour of iterative experiments for isolated trajectories. In case of the fast trajectories, further tuning of the parameters (e.g., learning rate or neural network size) may be required to improve the behavior. The results of the 6-fold cross validation experiments show that in most of the cases considered, the previously trained trajectories support the fast finding of pressure levels that are beneficial in terms of an improved energy-efficiency of the handling processes. In such cases where object losses occur due to mispredictions of the DQA, parameter adaptations are necessary. The agent acts too incautious in these cases, repeatedly causing too little holding forces and therefore object loss. With a more conservative DQA, energy savings of 50% are estimated to be realistic and can be achieved within a relatively short period of training time.

The introduced method offers great potential for industrial application. In the scope of this work, the object was not varied and the robot trajectory was analyzed manually. However, in future applications, it is straightforward to implement an automated detection of information such as package weight, carton type, and the planned robot trajectory, e.g., through scanning a QR code or an RFID chip on the package. The industrial applicability of the DQA could also be made more flexible if the DQA and the control system including sensors and valve were deployed in a decentral system. Independent of the present system setup (e.g., regardless of the exact robot control system and the corresponding programming interface) and adjacent processes, this decentral system could work completely self-sufficient. This would make it possible to collect data from different applications and use cases in order to make the DQA more versatile and more capable of generalization, if needed. Finally, the comprehensive and scalable application of such an intelligent unit can enable immense energy savings in a broad spectrum of industrial pick and place processes.

Author Contributions: Conceptualization, J.B. and F.G.; methodology, J.B. and F.G.; software, J.B.; data curation, F.G.; visualization, F.G.; writing—original draft preparation, F.G.; writing—review and editing, F.A. and K.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the German Federal Ministry of Economic Affairs and Energy, grant number 03ET1559B.

Supported by:



on the basis of a decision
by the German Bundestag

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: <https://lnk.tu-bs.de/sf7HRt> (accessed on 12 April 2021).

Acknowledgments: The authors thank Harald Kuolt at J. Schmalz GmbH for his support.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. eMarketer. Retail e-Commerce Sales Worldwide from 2014 to 2024 (in Billion U.S. Dollars). Statista. 2021. Available online: <https://www.statista.com/statistics/379046/worldwide-retail-e-commerce-sales/> (accessed on 12 April 2021).
2. Gabriel, F.; Fahning, M.; Meiners, J.; Dietrich, F.; Dröder, K. Modeling of vacuum grippers for the design of energy efficient vacuum-based handling processes. *Prod. Eng. Res. Dev.* **2020**. [CrossRef]
3. Gabriel, F.; Bobka, P.; Dröder, K. Model-Based Design of Energy-Efficient Vacuum-Based Handling Processes. *Procedia CIRP* **2020**, *93*, 538–543. [CrossRef]
4. Wolf, A.; Schunk, H. *Grippers in Motion: The Fascination of Automated Handling Tasks*, 1st ed.; Hanser: München, Germany, 2018; ISBN 978-1-56990-715-3.

5. Hesse, S. *Grundlagen der Handhabungstechnik*; Carl Hanser Verlag GmbH & Co. KG: München, Germany, 2016; ISBN 978-3-446-44432-4.
6. Kirsch, S.-M.; Welsch, F.; Schmidt, M.; Motzki, P.; Seelecke, S. Bistable SMA vacuum suction cup. In Proceedings of the Actuator 2018 16h International Conference on New Actuators, Bremen, Germany, 25–27 June 2018.
7. Motzki, P.; Kunze, J.; Holz, B.; York, A.; Seelecke, S. Adaptive and energy efficient SMA-based handling systems. In *SPIE Smart Structures and Materials + Nondestructive Evaluation and Health Monitoring, San Diego, California, United States, Sunday 8 March 2015*; Liao, W.-H., Ed.; SPIE: Bellingham, WA, USA, 2015; p. 943116.
8. Motzki, P.; Kunze, J.; York, A.; Seelecke, S. Energy-efficient SMA Vacuum Gripper System. In Proceedings of the Actuator 16-15th International Conference on New Actuators, Bremen, Germany, 13–15 June 2016.
9. Welsch, F.; Kirsch, S.-M.; Motzki, P.; Schmidt, M.; Seelecke, S. Vacuum Gripper System Based on Bistable SMA Actuation. In Proceedings of the ASME 2018 Conference on Smart Materials, Adaptive Structures and Intelligent Systems, San Antonio, TX, USA, 10–12 September 2018.
10. Kuolt, H.; Kampowski, T.; Poppinga, S.; Speck, T.; Moosavi, A.; Tautenhahn, R.; Weber, J.; Gabriel, F.; Pierri, E.; Dröder, K. Increase of energy efficiency in vacuum handling systems based on biomimetic principles. In *12th International Fluid Power Conference (12. IFK)*; Weber, J., Ed.; Dresdner Verein zur Förderung der Fluidtechnik e. V. Dresden: Dresden, Germany, 2020.
11. Zhakypov, Z.; Heremans, F.; Billard, A.; Paik, J. An Origami-Inspired Reconfigurable Suction Gripper for Picking Objects with Variable Shape and Size. *IEEE Robot. Autom. Lett.* **2018**, *3*, 2894–2901. [[CrossRef](#)]
12. Li, X.; Dong, L. Development and Analysis of an Electrically Activated Sucker for Handling Workpieces with Rough and Uneven Surfaces. *IEEE/ASME Trans. Mechatron.* **2016**, *21*, 1024–1034. [[CrossRef](#)]
13. Okuno, Y.; Shigemune, H.; Kuwajima, Y.; Maeda, S. Stretchable Suction Cup with Electroadhesion. *Adv. Mater. Technol.* **2018**, *26*, 1800304. [[CrossRef](#)]
14. Liu, J.; Tanaka, K.; Bao, L.M.; Yamaura, I. Analytical modelling of suction cups used for window-cleaning robots. *Vacuum* **2006**, *80*, 593–598. [[CrossRef](#)]
15. Bing-Shan, H.; Li-Wen, W.; Zhuang, F.; Yan-Zheng, Z. Bio-inspired miniature suction cups actuated by shape memory alloy. *Int. J. Adv. Robot. Syst.* **2009**, *6*, 29. [[CrossRef](#)]
16. Follador, M.; Tramacere, F.; Mazzolai, B. Dielectric elastomer actuators for octopus inspired suction cups. *Bioinspir. Biomim.* **2014**, *9*, 46002. [[CrossRef](#)] [[PubMed](#)]
17. Fischmann, C. Verfahren zur Bewertung von Greifern für Photovoltaik-Wafer. 2014. Available online: <https://elib.uni-stuttgart.de/handle/11682/4582> (accessed on 12 April 2021).
18. Bahr, B.; Li, Y.; Najafi, M. Design and suction cup analysis of a wall climbing robot. *Comput. Electr. Eng.* **1996**, *22*, 193–209. [[CrossRef](#)]
19. Mantriota, G. Optimal grasp of vacuum grippers with multiple suction cups. *Mech. Mach. Theory* **2007**, *42*, 18–33. [[CrossRef](#)]
20. Mantriota, G. Theoretical model of the grasp with vacuum gripper. *Mech. Mach. Theory* **2007**, *42*, 2–17. [[CrossRef](#)]
21. Mantriota, G.; Messina, A. Theoretical and experimental study of the performance of flat suction cups in the presence of tangential loads. *Mech. Mach. Theory* **2011**, *46*, 607–617. [[CrossRef](#)]
22. Radtke, M. Untersuchungen zur Dimensionierung von Sauggreifern. Ph.D. Thesis, Technische Universität Dresden, Dresden, Germany, 1992.
23. Karako, Y.; Moriya, T.; Abe, M.; Shimakawa, H.; Shirahori, S.; Saitoh, Y. A practical simulation method for pick-and-place with vacuum gripper. In Proceedings of the 56th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE), Kanazawa, Japan, 19–22 September 2017; pp. 1351–1356, ISBN 978-4-907764-57-9.
24. Becker, R. *Untersuchungen zum Kraftübertragungsverhalten von Vakuumgreifern*; Verl. Praxiswissen: Dortmund, Germany, 1993; ISBN 3-929443-17-1.
25. Horak, M.; Novotny, F. Numerical model of contact compliant gripping element with an object of handling. In Proceedings of the International Carpathian Control Conference ICCS, Malenovice, Czech Republic, 27–30 May 2002; pp. 691–696.
26. Liu, X.; Hammele, W. Die Entwicklung von Sauggreifern mit Hilfe der Finite-Elemente-Methode. *KGK Kautschuk Gummi Kunststoffe* **2002**, *10*, 530–534.
27. Valencia, A.J.; Idrovo, R.M.; Sappa, A.D.; Guingla, D.P.; Ochoa, D. A 3D vision based approach for optimal grasp of vacuum grippers. In Proceedings of the 2017 IEEE International Workshop of Electronics, Control, Measurement, Signals and their Application to Mechatronics (ECMSM), Donostia, San Sebastian, Spain, 24–26 May 2017; pp. 1–6, ISBN 978-1-5090-5582-1.
28. Mahler, J.; Matl, M.; Liu, X.; Li, A.; Gealy, D.; Goldberg, K. Dex-Net 3.0: Computing Robust Vacuum Suction Grasp Targets in Point Clouds Using a New Analytic Model and Deep Learning. In Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, Australia, 21–25 May 2018; pp. 1–8, ISBN 978-1-5386-3081-5.
29. Sdahl, M.; Kuhlentötter, B. CAGD—Computer Aided Gripper Design for a Flexible Gripping System. *Int. J. Adv. Robot. Syst.* **2005**, *2*, 15. [[CrossRef](#)]
30. Gabriel, F.; Römer, M.; Bobka, P.; Dröder, K. Model-based grasp planning for energy-efficient vacuum-based handling. *CIRP Ann.* **2021**, *1–4*. [[CrossRef](#)]

31. Fritz, F.; von Grabe, C.; Kuolt, H.; Murrenhoff, H. Benchmark of existing energy conversion efficiency definitions for pneumatic vacuum generators. In Proceedings of the Re-Engineering Manufacturing for Sustainability: Proceedings of the 20th CIRP International Conference on Life Cycle Engineering, Singapore, 17–19 April 2013; Nee, A.Y.C., Song, B., Ong, S.-K., Eds.; Springer: Singapore, 2013. ISBN 978-981-4451-47-5.
32. Fritz, F.; Haefele, S.; Traut, A.; Eckerle, M. Manufacturing of Optimized Venturi Nozzles Based on Technical-Economic Analysis. In Proceedings of the Re-Engineering Manufacturing for Sustainability: Proceedings of the 20th CIRP International Conference on Life Cycle Engineering, Singapore, 17–19 April 2013; Nee, A.Y.C., Song, B., Ong, S.-K., Eds.; Springer: Singapore, 2013. ISBN 978-981-4451-47-5.
33. Kuolt, H.; Gauß, J.; Schaaf, W.; Winter, A. Optimization of pneumatic vacuum generators—heading for energy-efficient handling processes. In Proceedings of the 10th International Fluid Power Conference, Dresden, Germany, 8–10 March 2016; Dresdner Verein zur Förderung der Fluidtechnik e.V.: Dresden, Germany, 2016.
34. Fontanelli, G.A.; Paduano, G.; Caccavale, R.; Arpentì, P.; Lippiello, V.; Villani, L.; Siciliano, B. A Reconfigurable Gripper for Robotic Autonomous Depalletizing in Supermarket Logistics. *IEEE Robot. Autom. Lett.* **2020**, *5*, 4612–4617. [[CrossRef](#)]
35. Tanaka, J.; Ogawa, A.; Nakamoto, H.; Sonoura, T.; Eto, H. Suction pad unit using a bellows pneumatic actuator as a support mechanism for an end effector of depalletizing robots. *Robomech. J.* **2020**, *7*. [[CrossRef](#)]
36. Schaffrath, R.; Jäger, E.; Winkler, G.; Doant, J.; Todtermuschke, M. Vacuum gripper without central compressed air supply. *Procedia CIRP* **2021**, *97*, 76–80. [[CrossRef](#)]
37. Shao, Q.; Hu, J.; Wang, W.; Fang, Y.; Liu, W.; Qi, J.; Ma, J. Suction Grasp Region Prediction using Self-supervised Learning for Object Picking in Dense Clutter. 2019. Available online: <https://arxiv.org/pdf/1904.07402> (accessed on 12 April 2021).
38. Mahler, J.; Goldberg, K. Learning deep policies for robot bin picking by simulating robust grasping sequences. In Proceedings of the 1st Annual Conference on Robot Learning, Mountain View, CA, USA, 13–15 November 2017.
39. Han, M.; Liu, W.; Pan, Z.; Xue, T.; Shao, Q.; Ma, J.; Wang, W. Object-Agnostic Suction Grasp Affordance Detection in Dense Cluster Using Self-Supervised Learning. 2019. Available online: <https://arxiv.org/pdf/1906.02995> (accessed on 12 April 2021).
40. Jiang, P.; Ishihara, Y.; Sugiyama, N.; Oaki, J.; Tokura, S.; Sugahara, A.; Ogawa, A. Depth Image-Based Deep Learning of Grasp Planning for Textureless Planar-Faced Objects in Vision-Guided Robotic Bin-Picking. *Sensors* **2020**, *20*, 706. [[CrossRef](#)] [[PubMed](#)]
41. Iriondo, A.; Lazkano, E.; Ansuategi, A. Affordance-Based Grasping Point Detection Using Graph Convolutional Networks for Industrial Bin-Picking Applications. *Sensors* **2021**, *21*, 816. [[CrossRef](#)]
42. OHLE, e.V. Open Hybrid LabFactory. Available online: <https://open-hybrid-labfactory.de/> (accessed on 19 May 2021).
43. Mohri, M.; Rostamizadeh, A.; Talwalkar, A. *Foundations of Machine Learning*, 2nd ed.; The MIT Press: Cambridge MA, USA, 2018; ISBN 9780262039406.
44. Edelkamp, S.; Schrödl, S. *Heuristic Search: Theory and Applications*; Morgan Kaufmann is an imprint of Elsevier: Amsterdam, The Netherlands; Boston, MA, USA; Heidelberg, Germany, 2012; ISBN 9780123725127.
45. Nguyen, H.; La, H. Review of Deep Reinforcement Learning for Robot Manipulation. In Proceedings of the 2019 Third IEEE International Conference on Robotic Computing (IRC), Naples, Italy, 25–27 February 2019; pp. 590–595, ISBN 978-1-5386-9245-5.