

Retraction

Retracted: Multifunctional Robot Grasping System Based on Deep Learning and Image Processing

Wireless Communications and Mobile Computing

Received 19 September 2023; Accepted 19 September 2023; Published 20 September 2023

Copyright © 2023 Wireless Communications and Mobile Computing. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] X. Zhang and K. Ye, "Multifunctional Robot Grasping System Based on Deep Learning and Image Processing," *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 9273474, 7 pages, 2022.

Research Article

Multifunctional Robot Grasping System Based on Deep Learning and Image Processing

XinYu Zhang ^{1,2} and Kai Ye¹

¹School of Automation Science and Technology, Xi'an Jiaotong University, Xi'an 710049, Shaanxi, China

²School of Information Engineering, Xi'an University, Xi'an 710065, Shaanxi, China

Correspondence should be addressed to XinYu Zhang; zhangxinyu@xawl.edu.cn

Received 14 February 2022; Accepted 7 March 2022; Published 31 March 2022

Academic Editor: Kalidoss Rajakani

Copyright © 2022 Xin Yu Zhang and Kai Ye. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper firstly introduces the general architecture of the multifunctional harvesting robot grasping system; then, deep learning is used to investigate the object target recognition, and a set of target detection algorithm based on convolutional neural network is implemented; then, image processing technology is used to realize the function of target object localization, which can guide the multifunctional harvesting robot to complete the picking of the target multifunctional. The experimental results show that the multifunctional harvesting robot has a small calculation error of multifunctional coordinates and has a strong multifunctional recognition and positioning capability.

1. Introduction

With the development of automation technology, robotics has grown by leaps and bounds, and a variety of intelligent robots are used in industrial, medical, educational, and agricultural applications. Robot grasping is an important function of robots in industrial scenarios, which is traditionally implemented based on manual teaching methods and 2D or 3D model matching to obtain grasping postures [1]. The former does not address the need to capture any object in any pose; the latter requires a large number of templates for different objects to be created in advance, and more than one template may be needed for the same object, making it laborious to build a template search library [2]. In order to solve this problem, the common method is to build grasping classifiers through machine learning algorithms for grasping pose planning, but these grasping classifiers often require human feature design, which is empirical and heuristic, and different features need to be designed for different grasping conditions, making human feature design difficult, inefficient, and impractical [3].

In recent years, with the increase in GPU computing power and the rapid development of deep learning technol-

ogy, deep neural networks driven by big data can learn the deep features corresponding to things or behaviours, especially for feature learning and expression of two-dimensional data, such as in image classification, object detection, semantic segmentation, and behavioural recognition research fields; the effect achieved by convolutional neural networks has far surpassed the traditional detection algorithms and even surpass human recognition in some areas [4]. The introduction of deep learning algorithms in robot grasping research, where convolutional neural networks are used to extract feature information of grasping poses in a hierarchical manner, can solve important challenges that previously required human-designed grasping features [5].

Furthermore, the use of convolutional neural networks allows for autonomous and efficient learning of feature representations of grasping poses from a large training dataset compared to inefficient human-designed features, resulting in improved performance of grasping pose detection algorithms [6].

In summary, the application of deep learning techniques to robot grasping pose detection algorithms not only eliminates the tedious work of building templates and human-

designed features but also allows for efficient grasping planning of target objects, which is of great value for research. In this study, we investigate how to identify and locate objects in a scene and design a well-structured convolutional neural network as a feature extractor for grasping poses.

2. Related Work

2.1. Object Detection. Current deep learning-based object detection methods can be categorised into two types depending on the implementation, Two-Stage object detection algorithms and One-Stage object detection algorithms. The R-CNN algorithm proposed by [7] outperforms the OverFeat end-to-end processing method proposed by [8] and improves the performance by about 50% compared to the traditional object detection algorithm. The YOLO algorithm proposed in [9] follows the regression-based One-Stage approach of the OverFeat algorithm to achieve true end-to-end object detection, and its improved version, YOLOv2 [10], can achieve a detection speed of 155 fps, but the YOLO-like algorithm also has shortcomings, as it is not rational for detecting small objects and objects with overlapping parts, and the algorithm's generalization capability and localization frame accuracy are insufficient. In [11], an SSD detection algorithm is proposed, which predicts object regions on the feature maps output by different convolutional layers, outputs discrete default box coordinates of multiple scales and proportions, and uses small convolutional kernels to predict the coordinate compensation values and category confidence of the candidate boxes.

2.2. Current Status of Robot Gripping Research. The application of machine vision technology to achieve robot grasping is to determine the relationship between the camera coordinate system and the robot base coordinate system, use the technology related to vision algorithms to detect the 3D poses of the target object, and implement the grasping action according to the detection results. In [12], a matching method is proposed for estimating the target pose, where a library of templates of the target object is created, its SIFT feature points and corresponding descriptors are detected, and the single-response matrix between the template and the target is derived by a matching algorithm. Reference [13] also used SIFT features to build a sparse 3D model as a template based on the feature point pairs and used a matching algorithm to identify the predicted target's pose information. Reference [14] proposed an object recognition and pose detection algorithm with real-time feedback for object colour and shape characteristics; however, the algorithm is not stable due to the effect of illumination and background. Reference [15] used a sliding window approach to generate multiple candidate grasping poses and then used a neural network to discriminate the optimal grasping pose, which is a less efficient method to obtain the optimal solution by traversing the search. In [16], a random sampling method was attempted to reduce the time-consuming candidate frame generation, but the final result was not significant. In [17], a cascaded convolutional neural network was proposed, with the first half applying an R-FCN network

to locate the target capture position and coarse capture angle classification prediction for RGB images and the second half refining the capture angle prediction by Angle-Net. A large amount of training data is obviously required to make the network model eventually achieve better results, as proposed in [18], to transform the positional detection problem into an angle classification problem by using the rotation angle as the label; however, labelling such a high-quality grasping angle dataset requires a lot of labour and resources. To simplify the annotation of the grasp angles, [19] proposed a grasp detection network GDN with both the image and the grasp angle as input and the output as the rating of that grasp angle, selecting the highest rating from 18 candidate grasp angles (divided into 18 such as 180°) as the detection result. This approach obviously simplifies the dataset production process at the expense of computational efficiency.

In order to improve the efficiency of dataset production, [20] proposed to conduct simulated grasping experiments in a simulation environment to collect the dataset, and their proposed Dex-Net2.0 system can achieve fast and accurate grasping of known objects through its own deep learning system for the 10,000 3D object models with different characteristics in the virtual object model library graph.

In an end-to-end implementation, [18] uses a neural network to generate the centroid of the grasp and the coordinates of the mechanical claw fingers directly from a local view of the target object. Reference [19], on the other hand, proposed a deep reinforcement learning algorithm for autonomous grasping based on deep Q-functions, whose experiments showed that the reinforcement learning algorithm could imitate complex movements and even plan the robot to achieve door opening.

3. Overall Design of the Multifunctional Harvesting Robot Gripping System

The studied multifunctional harvesting robot grasping system consists of an industrial camera, an image processing module, a multifunctional robot carrier, and a data processing unit (PC). The industrial camera is mounted on the robot arm of the multifunctional robot and is responsible for collecting information about the surrounding environment of the orchard and the multifunctional targets; the image processing module is responsible for processing the image information collected by the industrial camera; for example, the fruit robot is responsible for movement and grasping of the multifunctional targets; the data processing unit is responsible for further analysis of the image information to achieve identification and positioning of the multifunctional targets. The workflow of the multifunctional harvesting robot is shown in Figure 1.

When working, the industrial camera is used to capture images of the orchard environment and the multifunctional target, which is simply processed and compressed by the image processing module of the multifunctional robot and sent to the PC in the background; the PC in the background uses image processing and deep learning algorithms to preprocess, threshold segmentation, feature extraction, and camera calibration of the images; then, the module matches

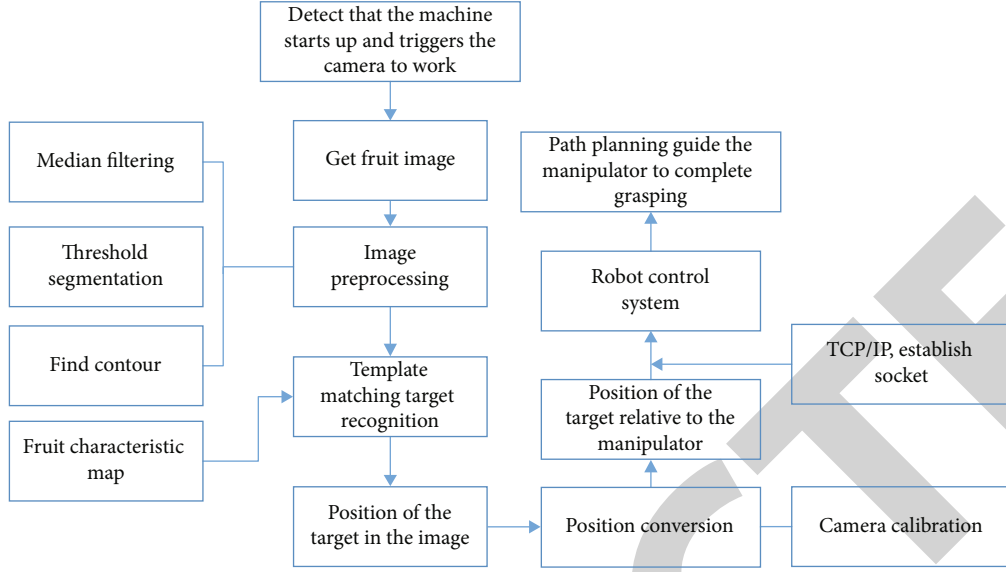


FIGURE 1: Workflow diagram of the multifunctional harvesting robot.

and identifies the position and coordinates of the multifunctional to be picked. After obtaining the relative position of the target multifunction relative to the multifunction robot, the background PC communicates with the multifunction robot via TCP/IP protocol, sends the coordinate information of the position of the multifunction to be picked to the robot, and guides the multifunction robot to carry out the picking operation to achieve the picking of the target multifunction.

4. Target Identification

The PC quickly identifies the multifunction, locates it in three dimensions, and sends the coordinates and picking instructions to the multifunction robot, which then completes the picking operation.

The former uses RCNN, FCN, and other algorithms to detect targets in the image region, while the latter transforms the detection problem into a regression problem to be solved. The latter is faster in detection but has lower accuracy and is not suitable for multifunctional robots with high accuracy in target recognition. Therefore, RCNN is used as the target detection algorithm for multifunctional robots.

The RCNN regional target detection algorithm differs from other algorithms in that it uses a convolutional neural network to extract the target features.

The RCNN regional target detection process consists of the following four steps:

- (1) Create a convolutional neural network model for extracting image features
- (2) Extract all the suspected regions in the image using a selective search method
- (3) Train a SVM classifier to classify the extracted features

- (4) Use regressor to correct the target regions to achieve target fruit recognition

Compared to traditional target feature extraction methods, the convolutional neural network structure used allows for the extraction and recognition of multilevel features; i.e., RCNN is able to ensure accurate recognition while avoiding complex manual feature extraction. The RCNN is also effective in detecting complex surroundings, as shown in Figure 2.

5. Target Localisation Based on Image Processing

After the backend PC identifies the multifunctional target, it also needs to obtain the location information of the target before the multifunctional robot can achieve the picking operation, which is very similar to the positioning method of the tennis picking robot. The image processing-based target positioning includes a backend PC and an industrial camera. The PC then uses the image processing algorithm to position the target multifunctionally based on the previous recognition results.

Based on the real coordinate system; O_{xy} and O_{uv} are the image and pixel coordinate systems, respectively; o_c is the industrial camera photocentre point; o is the image coordinate midpoint; P is the multifunctional real coordinate point of the target; p is the projection point of P in the image coordinate system; and f is the camera focal length. The relationship between the world and camera coordinate systems is obtained using rotation and translation, and the conversion equation is

$$T = (X_{OC-OW}, Y_{OC-OW}, Z_{OC-OW}), \quad (1)$$

where T is the translation matrix between the world and camera coordinate systems.

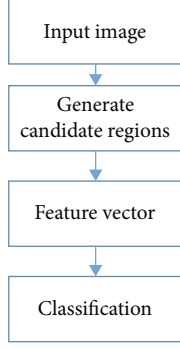


FIGURE 2: RCNN regional target detection flowchart.

The world coordinates $P(X_W, Y_W, Z_W)$ in the camera coordinate system $p(X_C, Y_C, Z_C)$ are

$$\begin{bmatrix} X_C \\ Y_C \\ Z_C \end{bmatrix} = R \begin{bmatrix} X_W \\ Y_W \\ Z_W \end{bmatrix} + T \Rightarrow \begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix} = \begin{bmatrix} R & R \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix}. \quad (2)$$

The coordinate transformation between camera and image can be derived from the projection principle as

$$\begin{aligned} x &= f \frac{X_C}{Z_C}, \\ y &= f \frac{Y_C}{Z_C}. \end{aligned} \quad (3)$$

Then, the coordinates $p_0(x, y)$ of point $p(X_C, Y_C, Z_C)$ in the camera coordinate system in the image coordinate system are

$$Z_C \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_C \\ Y_C \\ Z_C \end{bmatrix}. \quad (4)$$

The conversion expressions for the two are

$$\begin{aligned} u &= \frac{x}{dx} + u_0, \\ v &= \frac{y}{dy} + v_0 \end{aligned} \quad \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} 1/dx & 0 & u_0 \\ 0 & 1/dy & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}, \quad (5)$$

where dx and dy are the pixel equivalents and u_0 and v_0 are the image coordinate system origins.

TABLE 1: Coordinate data of the multifunctional robot gripping system.

	Fruit actual coordinates			Coordinate error		
	x	y	z	Δx	Δy	Δz
1	-0.593	-0.27	0.855	-0.2	-0.3	0
2	-12.357	-21.252	16.738	-0.1	-0.5	0.2
3	8.58	-5.264	8.625	0.3	0.2	0.1
4	15.714	8.521	0.859	-0.4	-0.1	0.2
5	-6.431	-1.371	23.083	-0.2	-0.5	0.1

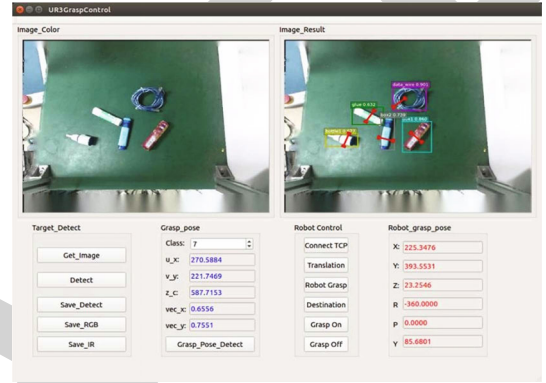


FIGURE 3: Robot gripping control software interface.

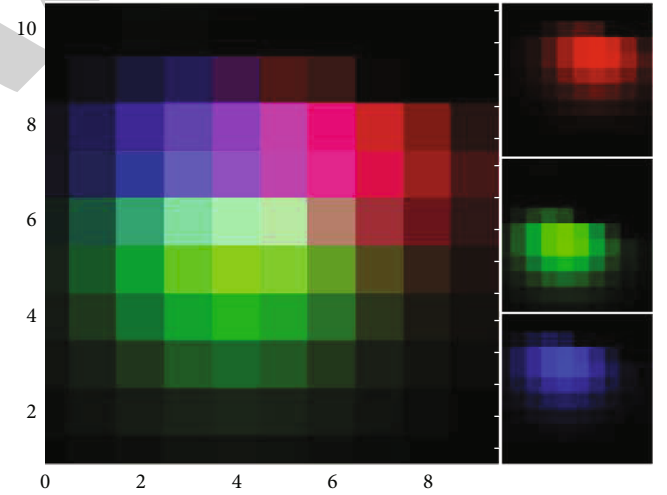


FIGURE 4: Target identification results.

6. Experimental Results and Analysis

In order to verify that the multifunctional harvesting robot grasping system based on deep learning and image processing meets the design requirements, practical grasping experiments were carried out using the system. In the grasping experiments, the accuracy of the multifunctional robot grasping system was judged mainly by the coordinate error of the target multifunction. The coordinate data of the multifunctional robot grasping system is shown in Table 1.

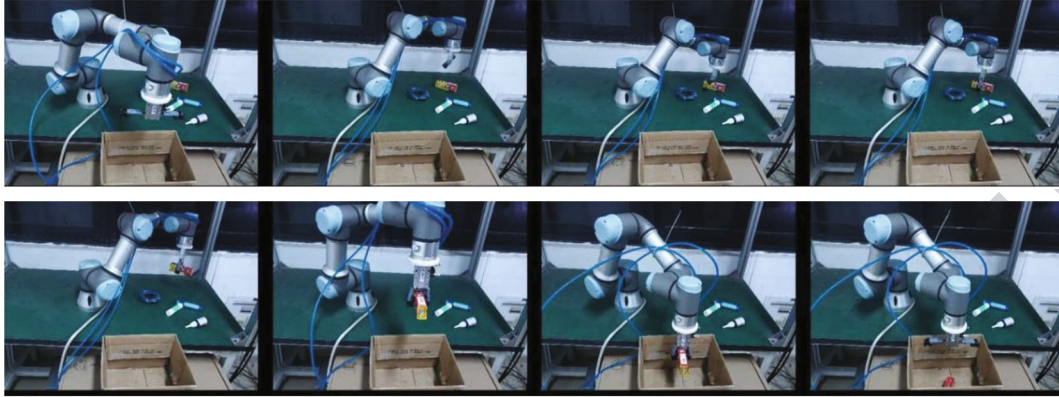


FIGURE 5: Flow diagram of the robot's autonomous gripping action.

Table 1 shows that the coordinate errors in all three axes are less than -0.5 , and in the z -axis, the errors are within 0.2 , indicating that the system has a strong multifunctional identification and positioning capability.

In order to further verify that the posture output of the proposed robot grasping posture detection system can achieve actual robotic two-finger grasping, an autonomous grasping experiment was set up to verify this. The human-computer interface is shown in Figure 3 and includes image display controls, function buttons, and parameter display controls. The Get_Image button controls the image acquisition and image preprocessing, while the Detect button implements the object detection function and displays the object detection result on the top right of the interface.

As shown in Figure 4, the target recognition effect can be achieved by adjusting the value of the Class control to display the grasping position parameters of the corresponding object; the Connect_TCP button enables communication between the computer and the robot console; the Translation button maps the grasping position to the robot grasping parameters and displays the parameter values; the Robot_Grasp button controls the movement of the robot to the corresponding position to grasp the object. The Robot_Grasp button controls the robot's movement to the corresponding position for grasping objects; the Grasp on and Grasp off buttons control the opening and closing of the gripper, respectively; the Destination button controls the robot's movement to the object placement point after the grasping action has been completed, and the robot grasps the object using this control software [21].

Only multiple target objects are placed in the detection area, as shown in Figure 5; for the flow chart of the robot to implement the grasping action, the specific grasping action flow is as follows:

- (a) The robot moves to a priming position with the air grasp in an open state
- (b) The robot control receives the grasping position of the target object, drives the robot to move directly above the target, and then adjusts the end joint angle to the yaw angle of the corresponding grasping position

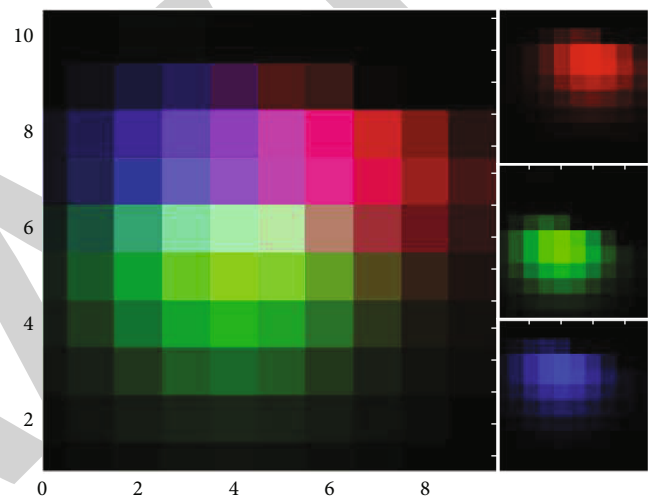


FIGURE 6: Optimised target recognition.

- (c) The robot moves downwards until it reaches the gripping position
- (d) The air grip closes and the gripping action is performed
- (e) The robot grips the target object and lifts it upwards
- (f) The robot transports the target object directly above the set placement point
- (g) The robot moves downwards to the object placement point
- (h) The air grip opens and releases the object, completing a single object gripping action

After a total of 100 actual gripping tests for each object tested, the overall gripping success rate was 84%. In the case of the metal bottle and the data cable, the results were not satisfactory, with a success rate of 60% for both items.

The main reason for the optimised target recognition as shown in Figure 6 is that the smooth metal bottle body affects the imaging effect of the depth map, which ultimately leads to a low success rate of effective grasping position

detection, while in the data line grasping experiment, the main reason for the low grasping success rate is limited by the hardware configuration of the two-finger gripper; the two-finger gripper used in the experiment is a pneumatic hand claw, which cannot be fully closed, and after closing, there is still a 16 mm gap between the two fingers. There was still a 16 mm gap between the two fingers, which did not allow for a secure grip on the small cross-sectional width of the cable, and the air grip failed to make a secure contact when closed. This is the main reason for the 7% difference between the overall gripping success rate and the effective gripping posture detection success rate, which could be improved by replacing the gripper with a better performing one.

The autonomous robot grasping experiments in this section demonstrate that the proposed grasping pose detection system is effective in the robot grasping planning task by mapping the pose results to the robot base coordinates for the target object.

7. Conclusions

Using an industrial camera as the image acquisition sensor, the multifunctional robot transmits the image information collected by the industrial camera to the PC in the background in real time through the network, and the PC uses image processing and deep learning algorithms to recognize and analyze the image, send the coordinates of the multifunctional to be picked to the multifunctional robot, and guide it to carry out the picking operation, which realizes the robot's grasping and positioning of the target multifunctional and is of great significance to improve the recognition accuracy of the multifunctional robot and realize the automation of multifunctional picking.

Data Availability

The datasets used in this paper are available from the corresponding author upon request.

Conflicts of Interest

The authors declared that they have no conflicts of interest regarding this work.

References

- [1] J. Kim, O. Nocentini, M. Scafuro et al., "An innovative automated robotic system based on deep learning approach for recycling objects," in *Proceedings of the 16th International Conference on Informatics in Control, Automation and Robotics*, pp. 613–622, Prague, Czech Republic, 2019.
- [2] W. SEBBATA, M. A. KENK, and J.-F. BRETHER, "An adaptive robotic grasping with a 2-finger gripper based on deep learning network," in *In 2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, pp. 620–627, Vienna, Austria, 2020.
- [3] G. Ghazaei, A. Alameer, P. Degenaar, G. Morgan, and K. Nazarpour, "Deep learning-based artificial vision for grasp classification in myoelectric hands," *Journal of Neural Engineering*, vol. 14, no. 3, article 036025, 2017.
- [4] C. Almanza, J. M. Baquero, and R. Jiménez-Moreno, "Robotic hex-nut sorting system with deep learning," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 4, pp. 3575–3583, 2021.
- [5] 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," *The International journal of robotics research*, vol. 37, no. 4-5, pp. 421–436, 2018.
- [6] J. Zhang, M. Li, Y. Feng, and C. Yang, "Robotic grasp detection based on image processing and random forest," *Multimedia Tools and Applications*, vol. 79, no. 3-4, pp. 2427–2446, 2020.
- [7] P. Jiang, Y. Ishihara, N. Sugiyama et al., "Depth image-based deep learning of grasp planning for textureless planar-faced objects in vision-guided robotic bin-picking," *Sensors*, vol. 20, no. 3, p. 706, 2020.
- [8] E. Johns, S. Leutenegger, and A. J. Davison, "Deep learning a grasp function for grasping under gripper pose uncertainty," in *In 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4461–4468, Daejeon, Korea (South), 2016.
- [9] M. Schwarz, A. Milan, C. Lenz et al., "NimbRo picking: versatile part handling for warehouse automation," in *In 2017 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3032–3039, Singapore, 2017.
- [10] K. Suzuki, M. Kanamura, Y. Suga, H. Mori, and T. Ogata, "In-air knotting of rope using dual-arm robot based on deep learning," in *In 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 6724–6731, Prague, Czech Republic, 2021.
- [11] C. Z. Di Wu, L. Ji, R. Ran, H. Wu, and Y. Xu, "Forest fire recognition based on feature extraction from multi-view images," *Traitement du Signal*, vol. 38, no. 3, pp. 775–783, 2021.
- [12] J. Varley, J. Weisz, J. Weiss, and P. Allen, "Generating multi-fingered robotic grasps via deep learning," in *In 2015 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pp. 4415–4420, Hamburg, Germany, 2015.
- [13] B. Kehoe, A. Matsukawa, S. Candido, J. Kuffner, and K. Goldberg, "Cloud-based robot grasping with the Google object recognition engine," in *In 2013 IEEE International Conference on Robotics and Automation*, pp. 4263–4270, Karlsruhe, Germany, 2013.
- [14] S. Caldera, A. Rassau, and D. Chai, "Review of deep learning methods in robotic grasp detection," *Multimodal Technologies and Interaction*, vol. 2, no. 3, p. 57, 2018.
- [15] R. Liu, F. Nageotte, P. Zanne, M. de Mathelin, and B. Dresp-Langley, "Deep reinforcement learning for the control of robotic manipulation: a focussed mini-review," *Robotics*, vol. 10, no. 1, p. 22, 2021.
- [16] Q. Jia, J. Cai, Z. Cao, Y. Wu, X. Zhao, and J. Yu, "Deep learning for object detection and grasping: a survey," in *In 2018 IEEE International Conference on Information and Automation (ICIA)*, pp. 427–432, Wuyishan, China, 2018.
- [17] J. Li, Z. Li, Y. Feng, Y. Liu, and G. Shi, "Development of a human-robot hybrid intelligent system based on brain teleoperation and deep learning SLAM," *IEEE Transactions on Automation Science and Engineering*, vol. 16, no. 4, pp. 1664–1674, 2019.
- [18] M. Alonso, A. Izaguirre, and M. Graña, "Current research trends in robot grasping and bin picking," in *In the 13th international conference on soft computing models in industrial and environmental applications*, pp. 367–376, Springer, Cham, 2018.

- [19] L. Pinto and A. Gupta, "Supersizing self-supervision: learning to grasp from 50k tries and 700 robot hours," in *In 2016 IEEE international conference on robotics and automation (ICRA)*, pp. 3406–3413, Stockholm, Sweden, 2016.
- [20] H. Kang, H. Zhou, X. Wang, and C. Chen, "Real-time fruit recognition and grasping estimation for robotic apple harvesting," *Sensors*, vol. 20, no. 19, p. 5670, 2020.
- [21] K. Kleeberger, R. Bormann, W. Kraus, and M. F. Huber, "A survey on learning-based robotic grasping," *Current Robotics Reports*, vol. 1, no. 4, pp. 239–249, 2020.

RETRACTED