

Vision Based Picking System for Automatic Express Package Dispatching

Shengfan Wang[†], Xin Jiang^{*}, Jie Zhao, Xiaoman Wang, Weiguo Zhou and Yunhui Liu, Fellow, IEEE

Abstract—This paper presents a vision based robotic system to handle the picking problem involved in automatic express package dispatching. By utilizing two RealSense RGB-D cameras and one UR10 industrial robot, package dispatching task which is usually done by human can be completed automatically. In order to determine grasp point for overlapped deformable objects, we improved the sampling algorithm proposed by the group in Berkeley to directly generate grasp candidate from depth images. For the purpose of package recognition, the deep network framework YOLO is integrated. We also designed a multi-modal robot hand composed of a two-fingered gripper and a vacuum suction cup to deal with different kinds of packages. All the technologies have been integrated in a work cell which simulates the practical conditions of an express package dispatching scenario. The proposed system is verified by experiments conducted for two typical express items.

I. INTRODUCTION

Recently, in many industry fields, human beings are replaced by robots. Although robot based automation has achieved great progress, there are still many tasks necessitate human labor. In logistics field, Amazon had held the robot competition (ARC) aiming to solve the problems involved in logistic process [1]–[3]. When Amazon first held the Amazon Picking Challenge in 2015, many teams failed to pick the specified items by their robotic systems. At that time, many teams only employed two fingered gripper as the end effector. While in the 2017 Amazon Robotics Challenge, most teams achieved high marks and employed both gripper and vacuum suction cup. It was proved in the challenge that the combination of gripper and vacuum suction cup was efficient. Although the challenge is terminated, it attract interests from the researchers in finding robotic solution in logistics.

One of the tasks in logistic system which is considered as a potential target for robotic automation is the process of package dispatching. Currently, this process is conducted by human workers. In the process, the worker have to pick up the package from the conveyor and recognize the destination information printed on it. Based on the information, the worker then dispatches the package to the specified line. The main

problems involved in the procedure is that the label printed on the package may not face upwards. In this case, the sensors can not capture the label and it needs the robot to grasp the package and then reverse it. In many situations, the packages are crowded randomly. This increases the difficulty in finishing the item grasping. In this research, we try to tackle the problems confronted in this situation by using RGB-D cameras.

II. RELATED WORK

For decades, researchers have been doing research on robot grasping [4]–[8]. In the early work [9], [10], human-designed features were used to represent grasps in images. In order to generate grasp candidate, full 3-D model of objects is necessary [11]. These methods was popular at that time, but they all faced challenges in the situations where no robust features nor full 3-D models are available. These situations are common in practical applications. Recently, many algorithms are proposed by to solve the robotic grasping problem without using CAD models. Andreas et al. [12], [13] proposed to directly generate grasp poses in point clouds and then use neural networks to rank all candidate poses in order to select the most proper one. Lenz et al. [14] attempted to employ deep learning technologies to learn and detect features representing proper robotic grasps using RGB-D images. In Amazon Robotics Challenge, team MIT-Princeton [15] built a multi-view vision system to estimate the 6D pose of objects. They also studied the policy of utilizing two functioned hand composed of two-fingered gripper and vacuum suction cup [16] Causo et al. [17] and Eppner et al. [2] designed robust robot systems for item picking integrated with perception, motion planning and special purpose end effector. Wade et al. [18] designed a multi-modal end-effector which is combined with three grasp synthesis algorithms.

The group in Berkeley did a lot of researches on robust grasp planning. They build a large dataset for grasp and suction planning which is called Dexterity Network. The dataset is composed of huge sensor data in various kinds of grasping scenarios and the corresponding metrics for evaluating grasp candidates. Both the grasping methods with parallel-jaw gripper and that using vacuum suction cup are considered in the dataset. They proposed the Grasp Quality Convolution Neural Network for evaluating grasp candidate directly from sensor input. This neural network is trained on the dataset of Dex-Net. The grasp and suction evaluation algorithm directly samples candidates from depth image with no object models assumed and this feature makes the method capable of dealing with novel objects.

This work was supported by the following projects: Shenzhen Peacock Plan Team grant (KQTD20140630150243062), Shenzhen and Hong Kong Joint Innovation Project (SGLH20161209145252406), Shenzhen Fundamental Research grant (JCYJ20170811155308088).

Shengfan Wang, Xin Jiang, Jie Zhao, Xiaoman Wang and Weiguo Zhou are with the School of Mechanical Engineering and Automation, Harbin Institute of Technology, Shenzhen 518055, China. The author e-mail: 18S053234@stu.hit.edu.cn. The corresponding author email: x.jiang@ieee.org.

Yunhui Liu is with the Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong, China.



Fig. 1. The express bag and envelope tackled in the dispatching procedure

Our work is inspired mainly by the algorithms proposed by the group in Berkeley and we have made the following contributions. Firstly, we improve the grasp sampling algorithm and it demonstrated better performance compared with the original one when dealing with the picking problems confronted with express package dispatching application. Secondly, we design a two-functioned robot hand consisting of a two-fingered gripper and a vacuum suction cup. Finally, by combining the methods for object detection YOLO [19], [20] with the Open Source Robot Operating System, we integrated a robot system shown in Fig. 2, with which a typical express package dispatching demonstration is realized.

III. PROBLEM STATEMENT

In the subsequent demonstration of package dispatching procedures, two kinds of items: bags and envelopes are considered. In practical dispatching line, they are processed separately. The respective conveyors transmit items to dispatching work cell. A worker has to recognize the information printed on the label of the package then dispatch them according to the information recognized. Since the status of the packages on the conveyor are random and overlapped with each other as shown in Fig. 1, the workers have to reverse the package if it does not face upward in order to see the label. The difficulties in a robotized solution for this procedure mainly come from the task of picking one package from overlapped ones. A successful dispatching requires that for all the packages the barcodes printed on them face upward and are recognized correctly.

IV. THE APPROACH

Our automatic express dispatching system is composed of two components, the vision processing system and the manipulator. As for the vision processing system, two RealSense D435 cameras are employed to provide color and depth images. The color images are used to detect objects and recognize the barcodes, while the depth images are fed to grasping planning algorithm. As for the manipulator, we use one UR10 robot mounted with an end-effector which support the usage of switching between using two-fingered gripper and a vacuum suction cup. All the programs of the system are implemented with Robot Operating System. Fig. 2 shows the whole system and the flowchart of the information processing pipeline is shown in Fig. 3.

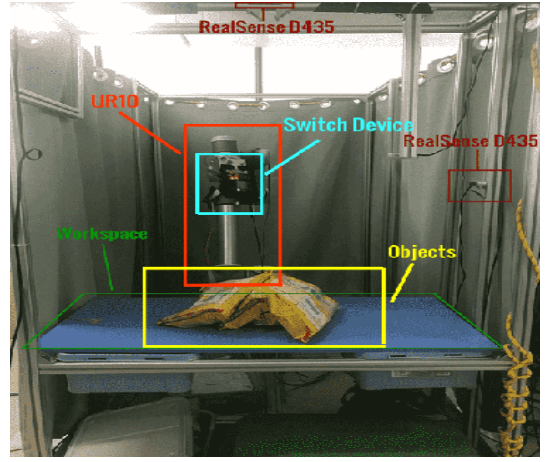


Fig. 2. The robot system

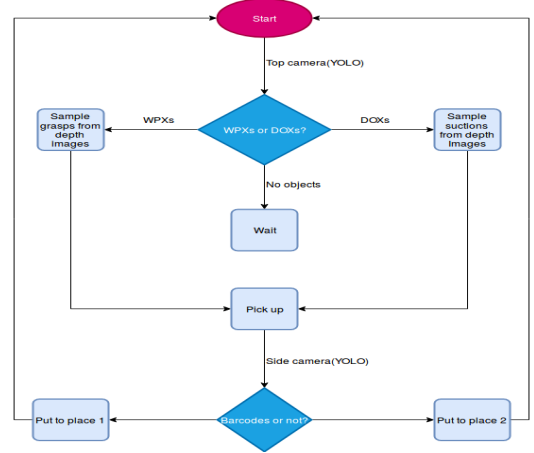


Fig. 3. The flowchart of the processing in the express package dispatching system

A. Package Recognition Method

The neural network frame YOLO is popular in object detection field for its accuracy and real-time performance when compared to Faster R-CNN [21]. For the purpose of package recognition, we tried both the official implementation of YOLO and Faster R-CNN based detection methods from Darknet and Google. They both perform well for our tasks. But there is a trade-off between the accuracy and speed. Faster R-CNN achieves higher accuracy while YOLO demonstrates faster speed. Considering the requirement of the real time performance in practical applications, we finally choose YOLO as the detector.

For the target of envelope and bag, we prepared for 25 pictures for each category respectively. Each picture contains multiple objects and all the objects are placed randomly like the situations of real industrial environment. We resize each picture to 640×480 and image argumentation is conducted to the dataset by random rotating each picture. It finally results to a dataset consisting of 300 images.

In order to detect express packages and barcodes, two



Fig. 4. Envelop and bag recognition



Fig. 5. Barcodes recognition

YOLO networks are trained separately. One of them is used for package detection and the other is used for barcode recognition. The processing pipeline is designed assuming that the robot first picks one item up using the camera configured above the workspace and then the other camera configured next to the workspace will be triggered to detect whether there is a barcode. Based on the recognition result, the robot will decide how to process the object. For those packages which barcodes face upward, the robot will place them on the conveyor directly, otherwise the object will be reversed to make its barcode face upward. After training of 20000 epochs, the networks achieves good performance in distinguishing envelopes from bags as shown in Fig. 4 and barcode recognition shown in Fig. 5.

B. Grasp Policy

Considering the deformable property of the express bags, we choose to use two-fingered gripper to grasp them. The grasp planning is implemented following the method which samples antipodal grasps directly from depth images [22]. After obtaining hundreds of candidate grasps, the GQ-CNN [22] is used to rank them and choose the best one. When we use the original grasp policy proposed by Berkeley, we find that it is likely to generate unreasonable results, which will lead to failure of grasping. In addition, we do not want the sampled grasp candidates to be on the surface of the bags, since it may lead to the collision between end-effector and items enclosed inside the bag, as shown in Fig. 6. Without information on the objects inside the bag, the grasp plan indicates danger. Thus we attempted to improve the algorithm by taking all above considerations into account, and its detail is shown in Algorithm 1.

The improved algorithm works like a filter and it ensures that only the reasonable grasp candidate will be left. The method is inspired by [12], [13]. The idea behind the constraints described in the algorithm is to make sure that there

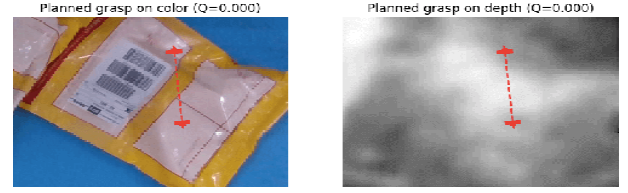


Fig. 6. The execution of the grasp may cause collision between the robot end-effector and the object enclosed inside the bag

Algorithm 1 Improved grasp sampling algorithm

Input: The set of candidate grasps generated by the original antipodal grasp sampling algorithm is denoted as \mathcal{G} . The depth value at the two parallel jaws position are represented separately as d_1 and d_2 . The depth value at the grasp center is denoted as d_0 . The color value at the two parallel jaws position are denoted separately, as c_1 and c_2 . The mean and standard deviation value of depth and color in the region covered by the grasp are denoted as $\mu_d, \mu_c, \delta_d, \delta_c$. The maximum and minimum depth value in the region covered by the grasp, are denoted as d_{\max} and d_{\min} . The parameters, ϵ in the following represents threshold values.

Output: The final grasp candidates set, $\tilde{\mathcal{G}}$,

```

1: for every  $g \in \mathcal{G}$  do
2:   if  $d_1 > d_0 + \epsilon_1$  and  $d_2 > d_0 + \epsilon_1$  then
3:     if  $d_{\max} - d_{\min} > \epsilon_2$  then
4:       if  $\mu_d > d_0 + \epsilon_3$  and  $\delta_d > \epsilon_4$  then
5:         if  $\delta_c > \epsilon_5$  then
6:           if  $\text{mean}(c_1 - c_2) > \epsilon_6$  then
7:              $g \cup \tilde{\mathcal{G}}$ 
8:           end if
9:         end if
10:       end if
11:     end if
12:   end if
13: end for
14: return  $\tilde{\mathcal{G}}$ 

```

is one part of the object in the grasp region as well as the color and depth distribution are different between the candidate grasp region and that of the outside. To be more specific, the depth and color value at the parallel jaw positions, grasp center positions are different with that outside of the parallel jaws. The optimal grasp for a bag with no information about the object enclosed in it is to grasp its corner. The proposed algorithm is to filter out the candidate satisfying the requirement.

Fig. 7 illustrates the whole sampling results from the improved algorithm. We could see that the algorithm outputs grasp candidates around the four corners of the bag. They are safer with low possibility of leading to collision between the robot hand and the objects inside the bag. The comparison between the proposed algorithm and the original one is demonstrated in Fig. 8. It is demonstrated that the proposed algorithm

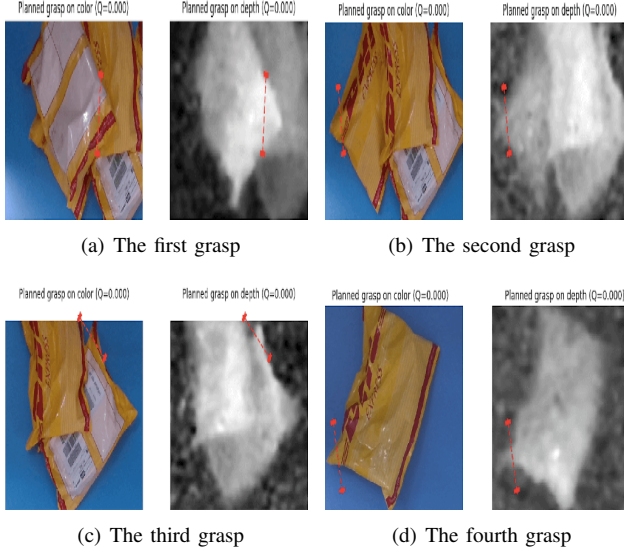


Fig. 7. The generated grasp plan during a whole subsequent picking process of four overlapped bags.

generate more reasonable grasp plan than that generated with the original one. We argue that the proposed method can also be applied to other similar situations.

C. Suction Policy

The envelopes are flat, which makes it suitable for using suction cup to directly suction on the surface in order to pick them up. We utilize the suction sampling policy proposed in the work [23]. We also assume that the position near the center of the envelope is the ideal place for suction. Therefore, we use YOLO as the detector to find where the envelope is and then take sampling suction point near center of the detection result. We use a vacuum cleaner to provide the necessary suction force. This kind of utilization is also popular in Amazon Picking Challenge [1]–[3], [15], [17], [18], [24], [25].

D. Gripper and Suction Cup Switch Device

In Amazon Picking Challenge, many teams employ a device to switch between gripping and suction mechanism in order to choose different picking ways according to the property of targets [1], [3], [18], [24], [25]. We also designed such a device to enable switch between a Robotiq 140 two-fingered gripper and Schmalz's suction as shown in Fig. 9.

Different from the other similar devices, our switch device also supports the usage of simultaneously employing both of the two mechanisms. In some situations, this provides merit for object manipulation. For example, for picking a book placed on a table, we can choose to firstly suction it. Since generally the width of the book will be beyond the that of the gripper jaws, it is impossible for gripping. After the book suctioned, we can then make the gripper to achieve a stable grasping.

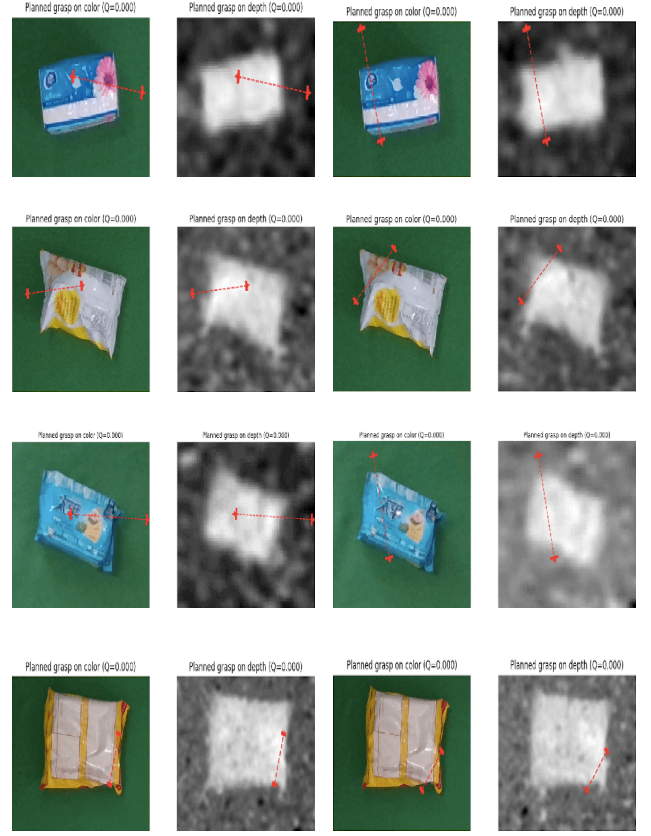


Fig. 8. Comparison between the original sampling algorithm(left) and the improved one (right)

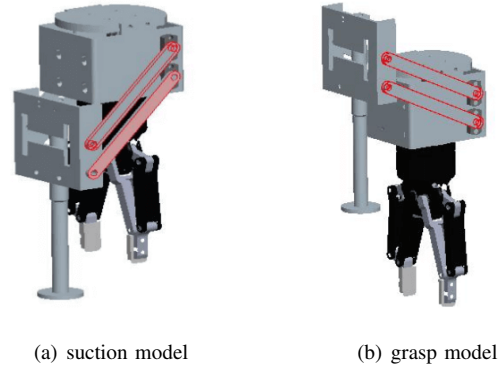


Fig. 9. Gripper and suction cup in two operation modes

V. EXPERIMENT

We implement all the proposed methods under ROS and conducted the corresponding experiments with our platform with a single GPU (NVIDIA GeForce GTX 1060). The experiment video can be viewed here ¹. In the verification experiments, we set the parameters in algorithm 1 as follows: $\epsilon_1 = 0.01$, $\epsilon_2 = 0.01$, $\epsilon_3 = 0.01$, $\epsilon_4 = 0.01$, $\epsilon_5 = 30$, $\epsilon_6 = 50$. Fig. 10 and Fig. 11 show the whole grasping and recognition

¹<https://youtu.be/TgD2G8B-QSY>



Fig. 10. Express bags picking experiment

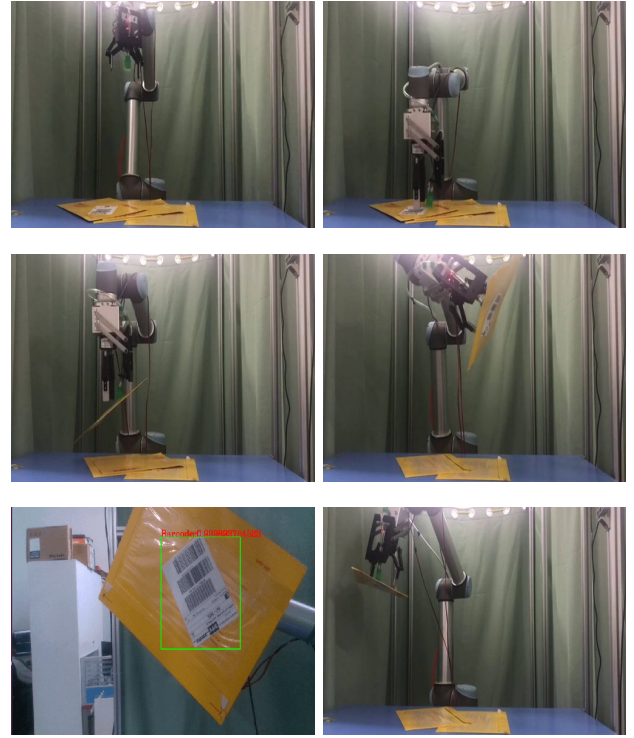


Fig. 11. Express envelopes picking experiment

procedure for express envelope and bag respectively. In each of the experiments, four objects are placed on the table in the initial state. The robot then conducted dispatching procedure of "pick-recognize-place" one by one. For the envelopes, the robot would choose to use suction and its success rate is almost 100%. For the bags, the robot would choose two-fingered gripper to complete the task. Its success rate is about 90%. The failures in the experiments are mainly due to the slip between the object and the gripper which leads to object dropping in transmitting phase. The time cost of picking and placing four bags is about fifty seconds and is about one minute and a half for four envelopes. For further study, we will try to reduce the time cost of whole process.

VI. CONCLUSION

In this research, we proposed a robot system to tackle the problem in express packages dispatching. In this procedure, a human worker has to pick up the packages and recognize the label on the packages. Then he should dispatch them according to the recognized information. The difficulty involved in the procedure is that the packages are randomly placed and overlapped with each other. In addition, without prior information on the objects enclosed inside the package, grasping may fail due to the contact between the robot hand and the object inside. For this problem an improved grasping planning method is proposed which generates grasp plan avoiding direct contact to the center of the package. For the purpose of package recognition, neural network based detection method is integrated. With the proposed methods, the robot system

demonstrated in experiments the capability of picking and recognizing two typical express packages: envelope and bag.

REFERENCES

- [1] N. Correll, K. E. Bekris, D. Berenson, O. Brock, A. Causo, K. Hauser, K. Okada, A. Rodriguez, J. M. Romano, and P. R. Wurman, "Analysis and observations from the first amazon picking challenge," *IEEE Transactions on Automation Science and Engineering*, vol. 15, no. 1, pp. 172–188, 2018.
- [2] C. Eppner, S. Höfer, R. Jonschkowski, R. Martín-Martín, A. Sieverling, V. Wall, and O. Brock, "Lessons from the amazon picking challenge: Four aspects of building robotic systems," in *Robotics: Science and Systems*, 2016.
- [3] 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 Transactions on Industrial Informatics*, 2018.
- [4] A. Bicchi and V. Kumar, "Robotic grasping and contact: A review," in *ICRA*, vol. 348. Citeseer, 2000, p. 353.
- [5] K. B. Shimoga, "Robot grasp synthesis algorithms: A survey," *The International Journal of Robotics Research*, vol. 15, no. 3, pp. 230–266, 1996.
- [6] A. Sahbani, S. El-Khoury, and P. Bidaud, "An overview of 3d object grasp synthesis algorithms," *Robotics and Autonomous Systems*, vol. 60, no. 3, pp. 326–336, 2012.
- [7] J. Bohg, A. Morales, T. Asfour, and D. Kragic, "Data-driven grasp synthesis survey," *IEEE Transactions on Robotics*, vol. 30, no. 2, pp. 289–309, 2014.
- [8] 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.
- [9] A. Saxena, J. Driemeyer, J. Kearns, and A. Y. Ng, "Robotic grasping of novel objects," in *Advances in neural information processing systems*, 2007, pp. 1209–1216.
- [10] A. Saxena, J. Driemeyer, and A. Y. Ng, "Robotic grasping of novel objects using vision," *The International Journal of Robotics Research*, vol. 27, no. 2, pp. 157–173, 2008.

- [11] A. T. Miller, S. Knoop, H. I. Christensen, and P. K. Allen, "Automatic grasp planning using shape primitives," in *Robotics and Automation, 2003. Proceedings. ICRA'03. IEEE International Conference on*, vol. 2. IEEE, 2003, pp. 1824–1829.
- [12] M. Gualtieri, A. ten Pas, K. Saenko, and R. Platt, "High precision grasp pose detection in dense clutter," in *Intelligent Robots and Systems (IROS), 2016 IEEE/RSJ International Conference on*. IEEE, 2016, pp. 598–605.
- [13] A. T. Pas, M. Gualtieri, K. Saenko, and R. Platt, "Grasp pose detection in point clouds," *International Journal of Robotics Research*, vol. 36, no. 13, p. 027836491773559, 2017.
- [14] I. Lenz, H. Lee, and A. Saxena, "Deep learning for detecting robotic grasps," *The International Journal of Robotics Research*, vol. 34, no. 4-5, pp. 705–724, 2015.
- [15] 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 *Robotics and Automation (ICRA), 2017 IEEE International Conference on*. IEEE, 2017, pp. 1386–1383.
- [16] A. Zeng, S. Song, K.-T. Yu, E. Donlon, F. R. Hogan, M. Bauza, D. Ma, O. Taylor, M. Liu, E. Romo *et al.*, "Robotic pick-and-place of novel objects in clutter with multi-affordance grasping and cross-domain image matching," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 1–8.
- [17] A. Causo, Z.-H. Chong, R. Luxman, Y. Yik Kok, Z. Yi, W. C. Pang, R. Meixuan, Y. Seng Teoh, W. Jing, H. Suratno Tju, and I.-M. Chen, "A robust robot design for item picking," 05 2018, pp. 7421–7426.
- [18] S. Wade-McCue, N. Kelly-Boxall, M. McTaggart, D. Morrison, A. W. Tow, J. Erskine, R. Grinover, A. Gurman, T. Hunn, D. Lee *et al.*, "Design of a multi-modal end-effector and grasping system: How integrated design helped win the amazon robotics challenge," *arXiv preprint arXiv:1710.01439*, 2017.
- [19] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.
- [20] J. Redmon and A. Farhadi, "Yolo9000: better, faster, stronger," *arXiv preprint*, 2017.
- [21] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91–99.
- [22] 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.
- [23] J. Mahler, M. Matl, X. Liu, A. Li, D. Gealy, and K. Goldberg, "Dex-net 3.0: Computing robust robot suction grasp targets in point clouds using a new analytic model and deep learning," *arXiv preprint arXiv:1709.06670*, 2017.
- [24] D. Morrison, A. W. Tow, M. McTaggart, R. Smith, N. Kelly-Boxall, S. Wade-McCue, J. Erskine, R. Grinover, A. Gurman, T. Hunn *et al.*, "Cartman: The low-cost cartesian manipulator that won the amazon robotics challenge," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 7757–7764.
- [25] K.-T. Yu, N. Fazeli, N. Chavan-Dafle, O. Taylor, E. Donlon, G. D. Lankenau, and A. Rodriguez, "A summary of team mit's approach to the amazon picking challenge 2015," *arXiv preprint arXiv:1604.03639*, 2016.