Learning-based Ellipse Detection for Robotic Grasps of Cylinders and Ellipsoids

Huixu Dong, Jiadong Zhou, Chen Qiu, Prasad K. Dilip, I-Ming Chen, *Fellow*, IEEE/ASME

Abstract*—In our daily life, there are many objects represented by cylindrical shapes and ellipsoids. The tops of these objects are formed by elliptic shape primitives. Thus, it is available for a robot to manipulate these objects by ellipse detection. In this work, we propose a novel approach to generating ground truth for training the model based on domain randomization. Using synthetic data generated in this manner, we build an end-to-end deep neural network with a detection backbone and then, combine multiple branches archived from the backbone for sharing the multiple-scale features; further, after employing active rotation filters, the features pass through the region proposal net to form the prediction branches of the box, orientation regression, and object classification; finally, these branches are fused to do ellipse detection, allowing robotic manipulations of cylinders and ellipsoids. To demonstrate the capabilities of the proposed detector, we show the comparison results with the state-of-the-art detector on synthetic and public datasets. The proposed model for ellipse detection and data generation pipeline based on domain randomization in a simulation are evaluated by a series of robotic manipulations implemented in real application scenarios. The results illustrate a high success rate on real-world grasp attempts despite having only been trained on a synthetic dataset. *(A video of some robotic experiments is available on YouTube: https://youtu.be/Ueg1XSI2S98)

***Index Terms*—Ellipse detection, Robotic grasp, Domain randomization.**

I. Introduction

AUOTOMATING industrial tasks, such as picking, binning, or assembly, the visual perception for robotic manipulation is essential [1-3]. It is well known that a large number of objects in households and industries have surfaces with elliptic geometric primitives. Therefore, the tops of cylindrical objects and the outlines of ellipsoid objects are represented by ellipses in the 2-dimensional(2D) images at most angles of observation, such as cans, cups, oranges and so on. Employing the detection of elliptic primitives, a robot can do static [4] and dynamic [2] manipulations of cylindrical objects.

The detection of ellipse-like shapes has been widely applied in various computer vision tasks, such as industrial inspection, medical diagnosis, recognition of traffic signs, tracking targets [5-8]. There exist a large number of ellipse detection works that are amenable to voting-based algorithms, algebraic analysis and geometric analysis of the properties of ellipses [2, 9]. Voting-based strategy includes methods [10, 11] that have reduced the dimensionality via voting in Hough Transform (HT) space [12, 13]. In terms of the algebraic approaches, a least-squares optimization problem [14] and random sample consensus [15] are in general applied to solving the ellipse fitting problem. Notably, the performances of these methods are more computationally efficient than HT-based approaches. Although the above methods have good performances in noisy images, they always generate many false positives for detecting multiple ellipses, which limits their applications in real environments. The second class contains the methods that employ geometric criteria to group arc segments with a high probability of belonging to the same ellipse [16]. In addition, some methods [2, 9, 13] utilized short straight lines to approximate arc segments for estimating elliptic parameters, which achieves a real-time faster and more accurate detection comparable with ones belonging to the first class. However, the above traditional ellipse-detection approaches highly rely on segmentation and grouping, which results in failure detections, especially in occluded and cluttered environments [17, 18], such as fruit detection and facial detection.

Figure. 1. The robot grasps a cylindrical object.

While some learning-based perception methods have been presented, the generalization of these models prevents them from being widely and easily applicable, specifically in robot-involved scenarios. The main challenge for the learning-based perception method used in real scenarios is data availability since labeling large amounts of training data

Huixu Dong (Corresponding author: dong0076@e.ntu.edu.sg), Jiadong Zhou, and I-Ming Chen are with Robotics Research Center of Nanyang Technological University, 639798 Singapore. Chen Qiu is with Maider Medical Industry Equipment Co., ltd, China 317607. Prasad K. Dilip is with Bio-AI Lab of Department of Computer Science at UiT The Arctic University of Norway, Tromso, Norway.

is generally time-consuming and needs expensive labor in the physical world. Recently some methods consider simulation as a tool to generate datasets for training models which are transferred successfully to the real world, bridging the simulation to the reality gap [19-23].

Herein, the goal of this work is to propose a network architecture of ellipse detection for accurately representing the elliptical objects for successfully inferring the whole information of each cylindrical or ellipsoid object for enabling a robot to manipulate such objects in real scenarios; see Fig. 1 for an example. In this work, the first step toward the goal of enabling a robot to grasp cylinders and ellipsoids by the ellipse detection is to improve the accuracy of the ellipse detection pipeline. Moreover, we investigate how to address the reality gap by creating the synthetic dataset based on domain randomization for sim-to-real transfer of cylinder and ellipsoid detection to realize robotic manipulation behaviors. Our main contributions are threefold and thus:

- An end-to-end one-stage model that detects cylinders and ellipsoids parameterized as ellipses is constructed. In particular, we first extract original features from the input image at three branches for the detection of objects with basic geometric primitives through the one-stage backbone [24], then combine multiple branches for sharing different-level features. Next, the features pass through the region proposal net after employing active rotation filters. Finally, the box regression, orientation regression and classification branches are fused to generate object detections.
- For better handling model training, we propose a pipeline of building synthetic datasets with labeled ellipses via the randomization domain. This method automatically samples transferred labels of ellipses by the constructed mathematical model for bridging the domain gap between the simulation and real scenarios. Allowing for domain randomization [19], we expose the 3D models of the cylindrical objects into various scenarios by Blenderproc [25] for extracting the ellipse ground truth. When the variability in simulation is significant enough, the proposed model will be generalized to the practical grasping environments.
- The perception strategy based on ellipse detection is implemented on an industrial robot for achieving a series of manipulations of cylindrical and ellipsoid objects in real scenarios, obtaining highly successful rates of grasp experiments.

We organize the rest of this paper as follows. The proposed model is described in detail in Section II. Section III investigates its performance by comparing it with a recent ellipse detector. Section IV provides demonstrations of robotic manipulations of cylindrical and ellipsoid objects. We conclude in Section V.

II. METHODOLOGY

This section consists of three main blocks, namely, data preparation by domain randomization, network architecture construction.

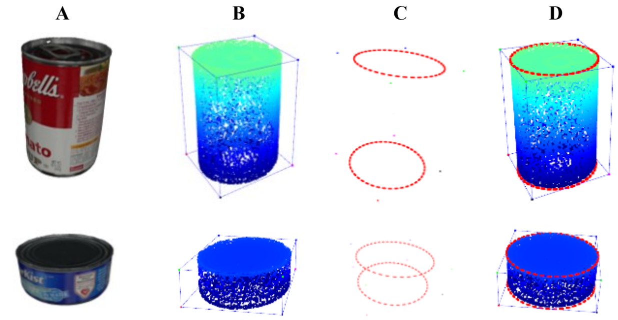

Figure. 2. The extraction of points corresponding to ellipses in the 3D space. The 3D models (A) and point clouds (B) of some cylindrical objects; The ellipse point sets extracted(C) and the combination visualization of 3D point clouds and ellipse point sets(D).

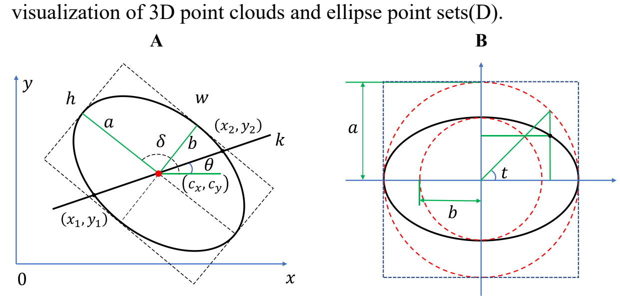

Figure. 3. The sketches of calculating the intersections of a line and an ellipse for extracting the point set consisting of an ellipse(A) and obtaining the bounding box covering the ellipse target(B). a, b, δ, c_x, c_y are the parameters determining an ellipse. k denotes a slope of a line. (x_1, y_1) and (x_2, y_2) represent the intersections of the line and ellipse. t is the parameterized angle for the parameterized ellipse function. δ indicates the orientation angle of an ellipse. θ is the orientation angle of the line.

A. Data Preparation by Domain Randomization

We present a novel approach to preparing a synthetic dataset including images with ellipses for training the proposed detector. Domain randomization technology is applied to providing enough simulated variability for bringing the gap from simulations to real scenarios. Here we create training images including cylindrical objects placed in scenarios and the air.

We render images using Blenderproc's built-in renderer by the two following methods. First, the object images are put in scenarios with textures that are chosen uniformly from 3D texture open libraries. Second, the images are randomly collected from the COCO dataset as the backgrounds of the object images in the air are composited to generate images. Moreover, we sample the tops of cylindrical or ellipsoid objects (such as apples and oranges) that are uniformly placed in scenarios and the air at random.

Here we present an approach to employing 3D point-cloud models of cylindrical objects to construct a synthetic dataset including images with ellipses. First, we extract the points consisting of ellipses that are the top outlines of cylinders after obtaining the 3D point-cloud models of cylinders (see Fig.3). In particular, an ellipse can be described by the following equation,

$$\frac{\left[(x-c_x)\cos\delta+(y-c_y)\sin\delta\right]^2}{a^2}+\frac{\left[(x-c_x)\sin\delta-(y-c_y)\cos\delta\right]^2}{b^2}=1 \quad (1)$$

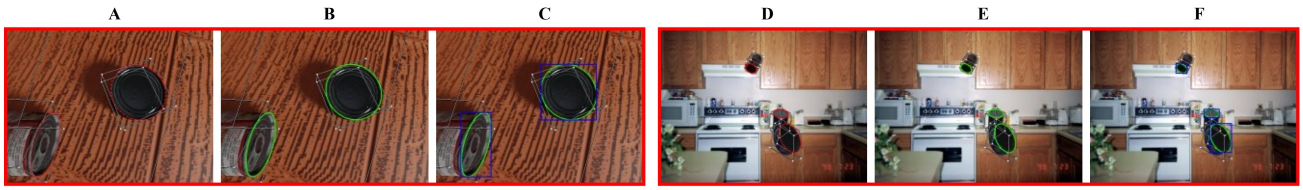

Figure. 4. The procedures of generating the ellipse ground truth with virtual 3D scenes(A, B, C) and real backgrounds(D, E, F). The point sets in red projected from 3D space(A and D); the ellipses in green fit from the point sets(B and E); the bounding boxes in blue covering the ellipses(C and F).

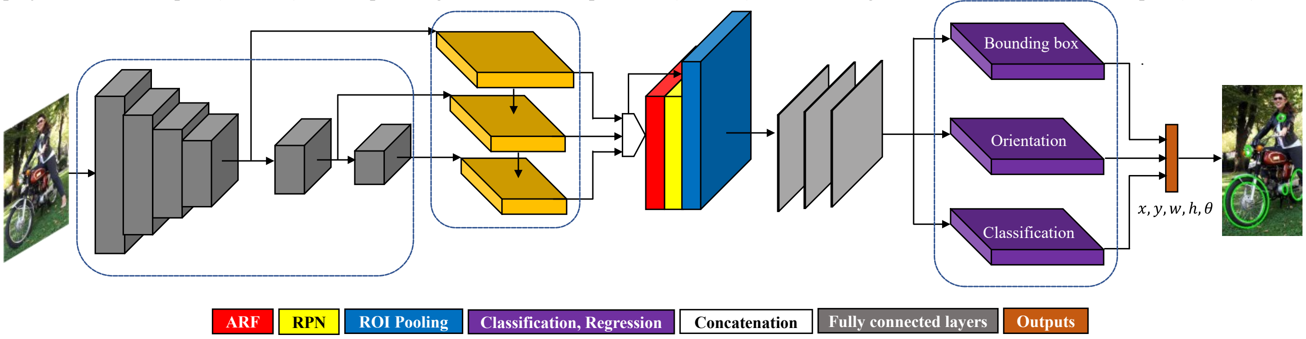

Figure. 5. The overview of the proposed detection model. The gray blocks enclosed by a frame represent the backbone and the yellow blocks enclosed by a frame denote the feature pyramid network.

$$y - c_y = k(x - c_x) \text{ with } k = \tan\theta \quad (2)$$

where (x, y) denotes the coordinate of a point on the ellipse; a, b, δ, c_x, c_y represent the semi-major, semi-minor axes, ellipse orientation angle, and the center coordinates, respectively. By combining Eq. (1) and Eq. (2), we can obtain the coordinates of the intersection points provided as

$$x_{1,2} = c_x \pm \frac{ab}{\sqrt{b^2(\cos\delta + k\sin\delta)^2 + a^2(\sin\delta - k\cos\delta)^2}} \quad (3)$$

$$y_{1,2} - c_y = k\left(x_{1,2} - c_x\right) \quad (4)$$

in which (x_1, y_1) and (x_2, y_2) denote the two coordinates of the intersection points, as shown in Fig. 2(A). Furthermore, by changing the ellipse orientation angle θ , we can achieve a series of intersection points (x_1, y_1) and (x_2, y_2) consisting of an ellipse, as shown in Fig. 2(C,D). Second, Blenderproc [25] is applied to rendering 3D models of objects to scenarios and generating 6D poses of objects. Third, we project the above intersection points of the object's point cloud from the 3D space to an image, via the 6D pose information provided by Blenderproc. The pinhole camera model is used for formulating this projection relationship. Fourth, the point sets in the 2D space projected from the 3D space are fit to an ellipse due to the geometric invariance of an ellipse, as illustrated in Fig. 4 (A, B, D, E). Finally, we need to determine the bounding box of the target ellipse for training. Equation (1) can be re-written as the parameterized equation,

$$\begin{aligned} x - c_x &= a\cos t\cos\delta - b\sin t\sin\delta \\ y - c_y &= a\cos t\sin\delta + b\sin t\cos\delta \end{aligned} \quad (5)$$

Thus, we have

$$\begin{aligned} \frac{\partial x}{\partial t} &= -a\sin t\cos\delta - b\cos t\sin\delta \\ \frac{\partial y}{\partial t} &= -a\sin t\sin\delta + b\cos t\cos\delta \end{aligned} \quad (6)$$

When $\frac{\partial x}{\partial t} = 0$ and $\frac{\partial y}{\partial t} = 0$, we can obtain

$$t = n\pi + \tan^{-1}\left(-\frac{b}{a}\tan\delta\right), n = 0,1$$

$$t = n\pi + \tan^{-1}\left(\frac{b}{a}\cot\delta\right), n = 0,1 \quad (7)$$

Substituting Eq. (7) into Eq. (5), we achieve the minimal and maximum values on the x -axis and y -axis, respectively, as shown in Fig. 3(B).

B. *Construction of the learning model*

1) *Network Architecture*

Figure 5 illustrates the proposed network that is made up of two sections including the backbone network ResNet101[26] and the designed output network for detecting ellipses. The backbone network takes charge of extracting the image features as the input to three branches aiming to complete three tasks.

We get the three feature maps generated by the backbone-ResNet101 feedforward network. In terms of the designed network, we achieve three-level resolution features for realizing multiple-scale object detection by combing the low, middle, and high output feature maps through the backbone. Since these feature maps are not of the same resolution, we need to normalize the feature maps before integrating them, such as up-sampling, undergoing 1×1 and 3×3 convolutional layers. During this stage, we still output three feature maps rather than one feature map sharing classification and regression to consist of a feature pyramid network[27] for generating more complementary semantic information.

The ARF-RPN regression [28, 29] is used for generating high-quality rotated region proposals. Different from standard CNN features, the oriented region proposal network, which consists of orientated response layers and the followed RPN and orientation pooling to generate rotation-invariant features, is proposed to achieve detection candidates with the orientations. Active rotation filters (ARF)[28] are applied to constructing orientated response layers for encoding the orientation information. An ARF is represented by a feature map with a canonical filter and its rotated clones, being a $k \times k \times N$ filter, where k denotes the kernel size and N represents the number of rotations. We employ ARF to the input $X(n)$ to produce the output Y_i the following equation,

$$Y_i = \sum_{n=0}^{N-1} F_{\theta_i}(n) \cdot X(n), \theta_i = i \cdot \frac{2\pi}{N}, i = 0, \cdots, N-1 \quad (8)$$

where F_{θ_i} is the θ_i-rotated version of F , $F_{\theta_i}(n)$ and $X(n)$ are the n-th orientation channel of F_{θ_i} and X, respectively.

After applying ARF, the model generates extra channels to incorporate richer rotation information and then, 2D region proposals are extracted and scored through 2D anchors to generate feature maps with orientation channels. Next, we used the ROI pooling on each feature map for cropping the features to get the maximum oriented proposal response.

An oriented bounding box is used for fitting an ellipse determined by five parameters. This becomes more challenging compared with using a standard bounding box since another variable is involved in the model. However, predicting all of the five variables together in one process will increase the burden of the shared CNN features, which potentially generates a negative effect on the results of individual variables. To relieve this issue, the structure above is activated to form three separate heads that include class classification, 2D box regression, orientation regression after passing fully connected layers. Finally, we achieve the final output to predict the location parameters, the class score of each rotated bounding box, respectively.

2) Loss construction

To determine an ellipse, we regress five offsets, (t_x, t_y, t_w, t_h) and t_θ from the two independent branches, respectively. These offset are defined as follows:

$$t_x = \frac{(x-x_a)}{w_a}, t_y = \frac{(y-y_a)}{h_a},$$
$$t_w = log\left(\frac{w}{w_a}\right), t_h = log\left(\frac{h}{h_a}\right),$$
$$t_\theta = \tan(\theta - \theta_a) \quad (9)$$

in which x, y, w, h and θ denote the center coordinates, width, height, and rotation angle of a rotated bounding box, respectively. (x, y, w, h, θ) and $(x_a, y_a, w_a, h_a, \theta_a)$ are respectively for the rotated predicted box and rotated anchor box. The coordinate (t_x, t_y) s parameterized as an offset, from the point (x_a, y_a), and it is normalized by (w_a, h_a). The ground-truth offsets $t^* = (t_x^*, t_y^*, t_w^*, t_h^*, t_\theta^*)$ are given as

$$t_x^* = \frac{(x^*-x_a)}{w_a}, t_y^* = \frac{(y^*-y_a)}{h_a},$$
$$t_w^* = log\left(\frac{w^*}{w_a}\right), t_h^* = log\left(\frac{h^*}{h_a}\right),$$
$$t_\theta^* = \tan(\theta^* - \theta_a) \quad (10)$$

where $(x^*, y^*, w^*, h^*, \theta^*)$ are respectively for the ground-truth box. Converting the values of the ground truth to the offset form $t^* = (t_x^*, t_y^*, t_w^*, t_h^*)$ and t_θ^* , we employ the smooth-L1 loss[30] for calculating the loss,

$$L_{reg}(t, t^*) = \sum_{i \in \{x,y,w,h\}} smooth - L1\,(t_i^* - t_i), \quad (11)$$
$$L_{reg}(t_\theta, t_\theta^*) = smooth - L1(t_\theta^* - t_\theta) \quad (12)$$

$L_{cls}(t_c, t_c^*)$ represents the class probability, which is provided as

TABLE I. Comparison results for four datasets, PD, P, R, T represent the public dataset. Precision, recall and time, respectively.

	Methods	F-measure	P	R	T(ms)
PD	Dong's	0.63	0.78	0.52	20
	Ours	0.73	0.80	0.67	18

Figure. 6. Examples of ellipse detection in the combined public dataset. Green ellipses indicate ground truth. Dong's detector results are represented by blue ellipses. The detected results based on the proposed detector are visualized by red ellipses.

$$L_{cls}(t_c, t_c^*) = -(1 - \hat{t}_c)^\gamma \log(\hat{t}_c), \quad (13)$$
$$\hat{t}_c = \begin{cases} t_c, & if\ t_c^* = 1 \\ 1 - t_c, & otherwise \end{cases}$$

where t_c^* represents the class label of the ground truth for an oriented anchor, which is defined as $t_c^* = 1$ if the sample is positive, $t_c^* = 0$ when the sample is negative. Here a sample is considered positive if the IoU_{ah}^{gt} between the anchor and any ground-truth, and their angular difference δ_{ah}^{gt} satisfy $IoU_{ah}^{gt} > 0.7$ and $\delta_{ah}^{gt} < \frac{\pi}{12}$, respectively; otherwise, this sample is defined as negative. t_c is the predicted probability for this sample being an ellipse. The factor $(1 - \hat{t}_c)^\gamma$ can adjust the loss of examples with large t_c^* ($\gamma = 2$). The summation of four losses is provided as:

$$L_t = w_1 L_{reg}(t, t^*) + w_2 L_{reg}(t_\theta, t_\theta^*) + w_3 L_{cls}(t_c, t_c^*) \quad (14)$$

We balance these terms by the weights w_1, w_2, w_3.

III. ALGORITHM PERFORMANCE

Our detector is compared with a recent ellipse detection, namely Dong's detector [2], on the performance, through the same evaluation metrics on the public datasets. Considering

Figure. 7. The workflow of robotic manipulation. The initial position(A); the grasping trajectories (B, D); grasping and releasing the object(C, E).

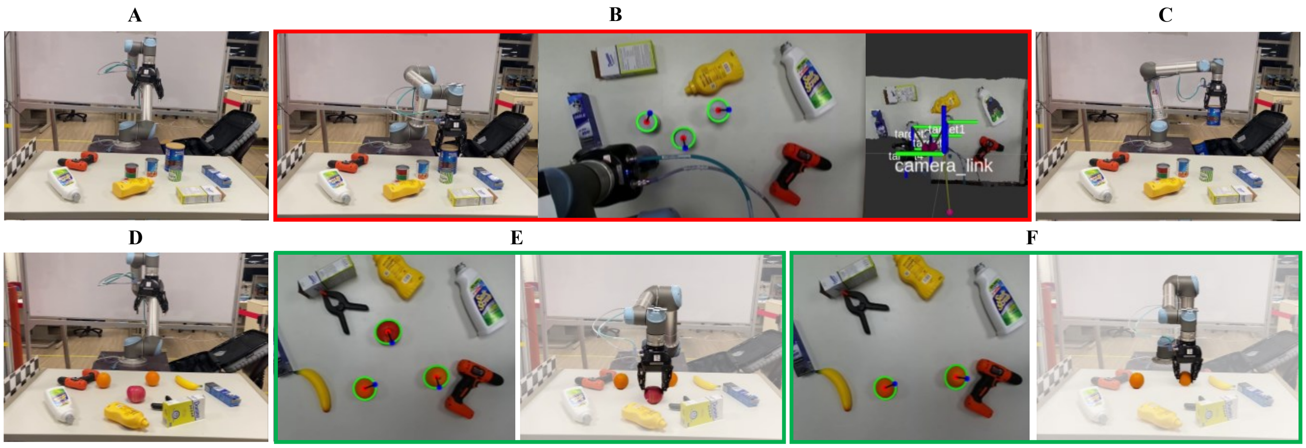

Figure. 8. The robot continuously grasps multiple cylindrical objects and ellipsoid objects in the cluttered scenario. Preparing to grasp(A, D) and releasing objects(C); the grasping scenario enclosed by a red frame is visualized in the front view (the first one), the top view (the second one) and Rviz view (the third one) (B); for each green frame, the first shows the detection results and the second illustrates the robot grasps the ellipsoid object(E, F).

the two main metrics, F-measure and the execution time factors together, our method outperforms Dong's method. We use the public datasets, including Dataset Prasad [13], dataset#1, dataset#2 [31] for evaluating the ellipse detection. The proposed detector is superior to Dong's method for F-measure, precision and recall, as illustrated in Table I.

Figure 6 illustrates several detected cases. The proposed detector and Dong's detector both cannot detect ellipses with very small size, such as Fig. 6(B), and have an unsatisfied performance on detecting some severely occluded ellipses, such as Fig. 6(A). The performance of our detector on detecting ellipse small ellipses is also better than that of Dong et al, as shown in Fig. 6(C). Dong's detector relies on combined arcs from three quadrants, leading to a valid detection when an ellipse consists of arcs from two quadrants due to the occlusion. Dong's method has a better performance than ours on fitting ellipse performance.

IV. Robotic Manipulation Experiments

A series of robot grasping experiments in real scenarios are conducted to evaluate the proposed perception method based on Sim2Real random domain through live video streams. The experimental setup consists of a 6-DOF industrial robot arm-Universal Robot 5 with a 3-finger Robotiq gripper. For capturing the targets, we mount a stereo camera ZED on a frame, as shown in Fig. 7.

A. Grasping cylindrical objects

Each experimental round, we fed some cylindrical objects frequently to a table. In each trial, the robot used the detection results in an RGB image to plan the grasping trajectory by continuously integrating the stream of incoming depth image for picking and delivering cylindrical objects, as illustrated in Fig. 8(A, B, C). We consider a successful grasp if the robot delivers a cylindrical object without dropping.

We repeated the pipeline for getting a statistical result to evaluate the detector's robustness. Out of the 50 total grasp attempts, 47 were successful resulting in a success rate of 94%. The proposed learning-based perception system can successfully detect cylindrical objects in the new scenarios. All the failures are attributed to the two following reasons. Since the images in the dataset used in the training phase do not reflect the real world perfectly, inaccurate detection cases occur. The remaining failure was that the gripper had one collision with an object as the robot almost arrived at a position beyond its operation space. The objects were pushed out of the workspace during grasp execution and therefore could not be delivered to the destination.

B. Packing fruits with 'ellipse' outlines

To illustrate that the proposed perception strategy can be used in agriculture scenarios, we conducted experiments of the robot grasping fruits whose outlines are similar to ellipses in cluttered environments. The multiple oranges were placed randomly on a table. The size of an orange is relatively small in a camera view. Moreover, there exists the calibration deviation in the grasping system and the used gripper does not include any sensor for providing grasping feedback. Thus, grasping could fail even with a slight error in perception. Our detection approach, however, overcame these issues and accurately detected multiple objects at the same time. Providing accurate perception feedback to the robot is the prerequisite of avoiding collision for realizing a successful grasp. We constructed a 'closed loop' grasp system. In particular, the robot continuously attempted multiple grasps until all targets were grasped, as illustrated in Fig. 8(D, E, F).

For quantitative results, we randomly placed some oranges and apples amidst clutter on a table. The robot first goes to a defined pre-grasp position, then executed a top-down grasp, yielding 3 trials per object. Overall, 48 fruits of a total of 50 fruits were successfully packed into boxes at first try except three ones that were ejected over by the gripper while moving a grasped object to the target container. Sources of grasping error include the ellipse detection algorithm and miscalibration between the camera and the robot.

V. Conclusion

In this paper, we constructed a learning-based ellipse detector for robotic grasps of cylinders and ellipsoids. A CNN network is proposed to detect ellipses with sufficient accuracy and speed to permit the robotic manipulation of cylinders and ellipsoids in complicated scenes. In grasping experiments, the robot successfully grasps cylinders and ellipsoid fruits in cluttered environments, which illustrates the proposed detector is potentially applicable to practical environments.

References

[1] H. Dong, D. K. Prasad, and I.-M. Chen, "Object Pose Estimation via Pruned Hough Forest With Combined Split Schemes for Robotic Grasp," IEEE Transactions on Automation Science and Engineering, 2020.

[2] H. Dong, E. Asadi, G. Sun, D. K. Prasad, and I.-M. Chen, "Real-time robotic manipulation of cylindrical objects in dynamic scenarios through elliptic shape primitives," IEEE Transactions on Robotics, vol. 35, no. 1, pp. 95-113, 2018.

[3] H. Dong, D. K. Prasad, Q. Yuan, J. Zhou, E. Asadi, and I.-M. Chen, "Efficient pose estimation from single RGB-D image via Hough forest with auto-context," in 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2018: IEEE, pp. 7201-7206.

[4] H. Dong, G. Sun, W.-C. Pang, E. Asadi, D. K. Prasad, and I.-M. Chen, "Fast ellipse detection via gradient information for robotic manipulation of cylindrical objects," IEEE Robotics and Automation Letters, vol. 3, no. 4, pp. 2754-2761, 2018.

[5] R. Jin, H. M. Owais, D. Lin, T. Song, and Y. Yuan, "Ellipse proposal and convolutional neural network discriminant for autonomous landing marker detection," Journal of Field Robotics, vol. 36, no. 1, pp. 6-16, 2019.

[6] R. S. Kothari, A. K. Chaudhary, R. J. Bailey, J. B. Pelz, and G. J. Diaz, "EllSeg: An Ellipse Segmentation Framework for Robust Gaze Tracking," IEEE Transactions on Visualization and Computer Graphics, vol. 27, no. 5, pp. 2757-2767, 2021.

[7] A. Fitriana, K. Mutijarsa, and W. Adiprawita, "Color-based segmentation and feature detection for ball and goal post on mobile soccer robot game field," in 2016 International Conference on Information Technology Systems and Innovation (ICITSI), 2016: IEEE, pp. 1-4.

[8] H. Dong, I.-M. Chen, and D. K. Prasad, "Robust ellipse detection via arc segmentation and classification," in 2017 IEEE International Conference on Image Processing (ICIP), 2017: IEEE, pp. 66-70.

[9] H. Dong, D. K. Prasad, and I.-M. Chen, "Accurate detection of ellipses with false detection control at video rates using a gradient analysis," Pattern Recognition, vol. 81, pp. 112-130, 2018.

[10] L. Xu, E. Oja, and P. Kultanen, "A new curve detection method: randomized Hough transform (RHT)," Pattern recognition letters, vol. 11, no. 5, pp. 331-338, 1990.

[11] N. Kiryati, Y. Eldar, and A. M. Bruckstein, "A probabilistic Hough transform," Pattern recognition, vol. 24, no. 4, pp. 303-316, 1991.

[12] H. I. Cakir, B. Benligiray, and C. Topal, "Combining feature-based and model-based approaches for robust ellipse detection," in Signal Processing Conference (EUSIPCO), 2016 24th European, 2016: IEEE, pp. 2430-2434.

[13] D. K. Prasad, M. K. Leung, and S.-Y. Cho, "Edge curvature and convexity based ellipse detection method," Pattern Recognition, vol. 45, no. 9, pp. 3204-3221, 2012.

[14] D. K. Prasad, M. K. Leung, and C. Quek, "ElliFit: An unconstrained, non-iterative, least squares based geometric Ellipse Fitting method," Pattern Recognition, vol. 46, no. 5, pp. 1449-1465, 2013.

[15] F. Mai, Y. Hung, H. Zhong, and W. Sze, "A hierarchical approach for fast and robust ellipse extraction," Pattern Recognition, vol. 41, no. 8, pp. 2512-2524, 2008.

[16] Q. Ji and R. M. Haralick, "A statistically efficient method for ellipse detection," in Image Processing, 1999. ICIP 99. Proceedings. 1999 International Conference on, 1999, vol. 2: IEEE, pp. 730-734.

[17] P. Roy and V. Isler, "Vision-based apple counting and yield estimation," in International Symposium on Experimental Robotics, 2016: Springer, pp. 478-487.

[18] W. Dong, P. Roy, C. Peng, and V. Isler, "Ellipse r-cnn: Learning to infer elliptical object from clustering and occlusion," IEEE Transactions on Image Processing, vol. 30, pp. 2193-2206, 2021.

[19] J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, and P. Abbeel, "Domain randomization for transferring deep neural networks from simulation to the real world," in 2017 IEEE/RSJ international conference on intelligent robots and systems (IROS), 2017: IEEE, pp. 23-30.

[20] K. Bousmalis et al., "Using simulation and domain adaptation to improve efficiency of deep robotic grasping," in 2018 IEEE international conference on robotics and automation (ICRA), 2018: IEEE, pp. 4243-4250.

[21] G. Shi et al., "Fast Uncertainty Quantification for Deep Object Pose Estimation," arXiv preprint arXiv:2011.07748, 2020.

[22] J. Tremblay, S. Tyree, T. Mosier, and S. Birchfield, "Indirect object-to-robot pose estimation from an external monocular rgb camera," in 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2020: IEEE, pp. 4227-4234.

[23] J. Tremblay, T. To, B. Sundaralingam, Y. Xiang, D. Fox, and S. Birchfield, "Deep object pose estimation for semantic robotic grasping of household objects," arXiv preprint arXiv:1809.10790, 2018.

[24] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," arXiv preprint arXiv:1804.02767, 2018.

[25] M. Denninger et al., "Blenderproc," arXiv preprint arXiv:1911.01911, 2019.

[26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770-778.

[27] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 2117-2125.

[28] Y. Zhou, Q. Ye, Q. Qiu, and J. Jiao, "Oriented response networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 519-528.

[29] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," IEEE transactions on pattern analysis and machine intelligence, vol. 39, no. 6, pp. 1137-1149, 2016.

[30] R. Girshick, "Fast r-cnn," in Proceedings of the IEEE international conference on computer vision, 2015, pp. 1440-1448.

[31] M. Fornaciari, A. Prati, and R. Cucchiara, "A fast and effective ellipse detector for embedded vision applications," Pattern Recognition, vol. 47, no. 11, pp. 3693-3708, 2014.