ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 8| Issue: 7| July 2022|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2022: 8.205 || ISI Value: 1.188

DEEP CONVOLUTIONAL NEURAL NETWORK DESIGN APPROACH FOR 3D OBJECT DETECTION FOR ROBOTICGRASPING

[1]Mrs.Anitha J, [2]Mrs. Kapu Anithalaksmi

[1]Asst. Professor of Dr. Ambedkar Institute of Technology, Dept. of MCA, Bangalore – 560060, Karnataka, India
[2]Student of Dr. Ambedkar Institute of Technology, Dept. of MCA, Bangalore – 560060, Karnataka, India

ABSTRACT

Automation is increasing with the advent of 3Danddepth images using an infrared camera. Formerly, most of the object detection and recognition were done by a 2D-camera of Red-Green-Blue (RGB) images. Today, with the availability of economical 3D-sensors, people started investing in 3D-object detection. Over recent years, Convolutional Neural Network (CNN) has reached the epitome of image classification for different application. Explicitly in 2D-CNN, there is massive progress for object detection, but for 3D-CNN, it is still at the beginning of an era. In this proposal, RGB images and depth images will be used to classify the objects using a Structure Sensor camera. Previously, researchers mainly worked on 2D-images and received considerable accuracy, but the accuracy decreases for 3D-image classifications using 2D-techniques. The complexity of the proposed model is increased due to the partial images with variation in illumination, occlusion, and cluttered images. To overcome the complexity problem, Deep Convolutional Neural Network (DCNN)is proposed with the usage of RGB-D images, which includes one 2D-image in RGB, and the other in-depth image form. Two methods are proposed, one will be using two parallel DCNN model and another method will be using three parallel DCNN model. Afterwards, the parallel models need to be concatenated to get 3D- object detection. Public-available 3Ddataset will be used to evaluatethe model, and the results of both of the models will be compared.

KEYWORDS—***Deep Convolutional Neural Network, MachineLearning, 3D object recognition, detection.***

I. INTRODUCTION

With the understanding of 2D-images, 3D-understanding is also promoted from machine learning. Extracting essential features from 3D-images is a perplexing task in computer vision. Applications of computer vision ranging from mapping, 3Dsensing, robotics, augmented reality, autonomous driving demand of 3D-feature extraction is in high demand. Many fields, giant companies, academia, and industries are contributing to the 3D-computer vision. The main cause of rapid improvement of 3D-vision is its application and necessity in industries [1].

Assessment, image elucidation and feature extraction of significant features are natural and instantaneous for human. whereas for the robotic vision or computer vision finding the segmentation, orientation, pose, detection are profound problems and ambition for machine learning community [2]. Researchers are trying to imitate the vision ability of human being using high resolution with depth information cameras, and convolutional neural networks. In past decade, CNNs have become a very popular and powerful tool for the suspension of computer vision problems. However due to the complexities of 3D-robot-based vision, computer vision is still in their early infancy [3].

This work addresses the design approach in how to improve 3D object detection for a robot arm by utilizing 2D machine learning methods. The design methods consider the 3D parameters of the images into 2D sets of information, and process the object detection to obtain high accuracy performance of identification and distance information for the navigation and grasp of the robot arm.

In this proposed research, the ABB (ASEA Brown Boveri) company manufactures the industrial robot arm with its controller for the movement control. The robot arm is providedwith a gripper mounted on it to grasp the objects that will be detected by the robotic vision using the Structure Sensor/Core camera. An implementation of a vision system for detection and recognition of 3D-objects in RGB-D (Red-Green-Blue- Depth) image in point cloud form is captured by the StructureSensor/Core camera. Experimentation work will be performed under different illumination and challenging conditions such as cluttered, occluded, and obstructed partial images. Theproblem of detecting an object and recognizing it with the given 3D-point cloud segments includes cluttered background, occlusion, partial images, and low illumination. Most of the current stateof-art in this problem are using old classifiers, low- resolution camera, and basic deep learning CNN architecture, which provides a lower accuracy.

The Deep Convolutional Neural Network (DCNN) method is introduced in this proposal design in a Multi-Modal approach, where two different DCNNs are proposed for RGB and Depth for object detection. These images will be fed to the DCNN architecture after an image is pre-processing using OpenCV or Scikit image. Image pre-processing stage may be optional as it

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 8| Issue: 7| July 2022|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2022: 8.205 || ISI Value: 1.188

varies according to accuracy and results obtained from the experiments. In the DCNN architecture, features will be extracted using convolutional layers and for sub-sampling of the image pooling layer will be used. The output of these two layers will be merged, and then it will be fed to another convolutional layer for classification and detection purpose. In the proposed design, another method is discussed in where a 3D-image is split into three different planes and be processed with three different parallel DCNNs. The three different layers will be merged and connected to fully-connected layer for classification and detection. Further, optimization will be done using TensorFlow Inference to make computation faster. After these processes, object coordinates will be given to the robot arm controller to perform the grasping task. Finally, results of these methods will be compared for better and accurate performance. Contribution to the robotic vision community is:

BACKGROUND

Humans need assistance from robots, cobots (collaborative robots) or assistive robots, due to ever increasing demand of precision and conscientiousness of semiconductor industries, mobile phone manufacturing plants, automation industries, laser treatment using the robot in the health center, in the segregation of materials in UPS, FedEx, DHL, USPS, and many more. Robots can be found anywhere from industries, laboratories, and in offices. Recently, an article read about a humanoid named "Sophia" made by a Hongkong company named Hanson robotics, which made its first public appearance in Austin, Texas in 2016 [4]. Sophia's eyes combined with computer vision, so there must be an object detection, image processing in all conditions. This project requires computer vision to pick the previously specified items with high certainty in accuracy of its object detection capabilities.

Object detection and recognition idea leaped into notice after the Amazon Picking Challenge was initiated. The Amazon Picking Challenge allowed an individual to create their robotic system to help with mass production and management. It was first held at the 2015 International Conference on Robotics and Automation. It started to provide a challenge to the robotics research community that involved amalgamating the state of the art in objection and perception, motion planning, clasp planning, and chore planning. They constituted a simplified model of the task that many humans face in warehouses. It was supposed to pick items from the shelves and put them into receptacles [5].

The proposed method is concentrated more on warehouse and industrial applications. In warehouse environments, a human needs to deal with several different objects, which can be handled easily by the robot. Regardless of noteworthy accomplishments, the problem of detection and recognition of objects efficiently and accurately still remains a scientific challenge when real scenes are considered [7]. This project work proposes a design approach on 3D-image in the cluttered and occluded image with poor illumination. For warehouses, there are amorphous objects, different shapes, and structure.
Furthermore, blur, noisy images, complex shapes make it more complicated for identification and grasp of an object when challenges are presented to the image. To overcome this problem, some optimizing tools like TensorFlow Inference will be used in Python programming language for the design development.

In the proposed method, ABB IRB 120 is an Industrial Robot Arm manufactured by ABB that will be used which is shown in the Fig. 1. It is the fourth generation of robotic technology with 6-axis. It is perfect for part handling and assembly applications in electronic, manufacturing and food industry. It can handle a load of 3kg and is enabled with fast acceleration and can deliver accuracy for any application, which will help in the proposed design. To control the arm, it needs a robot controller, here IRC5 (Industrial Robot Controller) will be used. It gives the ability to perform the task in an efficient manner providing a pinpoint path with accuracy. IRC5 is programmable with a high-level programminglanguage. To grasp the object impactive gripper will be used which is mounted on robot arm.

Fig. 1. ABB manufacturing robot arm

II. DCNN DESIGN APPROACHES

The main goal of the proposed method is to design a model to successfully recognize the object in the warehouse under the partial view, poor lighting conditions, low visibility, andchallenging conditions. Furthermore, this model will be integrated with the ABB robot which can grasp the object from the warehouse shelves. In robotic vision, most of the researchers implement different variations of DCNNs for 3Dobject recognition and detection to get the optimized results for different objects in

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 8| Issue: 7| July 2022|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2022: 8.205 || ISI Value: 1.188

diverse challenging conditions. In this section, methods to be used in the research work is explained inthe brief and shown in Fig. 2.

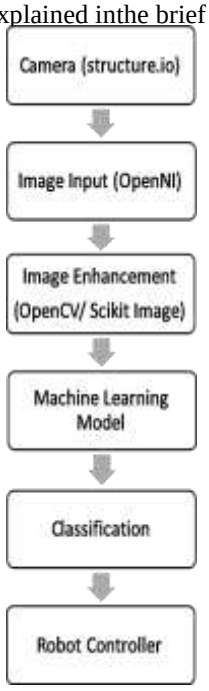

Fig. 2. Project methodology process

A. Dataset Acquisition

In this proposed research work, the main goal is to identify objects in a semi-structured environment that includes poor illumination, cluttered objects, occluded objects by other objects, low light, reflective surface, amorphous body, and partial views of an object. To train the proposed model, large number of samples in the dataset are needed containing several objects in warehouse environments. Through literature review, some datasets used for the APC competition and other purposes include blurry images, reflective images, low lights, high brightness, occluded, cluttered, partial, and RGB-D images of multiple diverse objects. In the proposed design, it is planned to use some public repository. The available datasets are particularly designed for pick-and-place robot and for 3Dgrasping purpose such as the CURE OR (Challenging Unreal and Real Environment for Object Recognition) [8] dataset, MIT Princeton dataset [9][10], and the RGB-D Washington dataset [11].

The RGB-D Washington dataset developed by researchers at Washington University and consists of 11,427 manually segmented RGB-D images. The images contain 300 common household objects which have be classified into 51 classes arranged using WordNet hypernym-hyponym relationships similar to ImageNet. The images are captured using a Kinect style sensor to generate 640x480 RGB-D frames and a frequency of 30 Hz [1]. These datasets contain most of the challenging conditions which is mentioned in the proposal. For the proposedmethod, it is planned to use the mentioned datasets including RGB-D images and in the point cloud form images captured. Point cloud is a set of 3D-data points, where each point is represented by three coordinates in a cartesian or another suitablecoordinate system. The image capture from the Structure Sensor/Core camera requires the utilization of the OpenNI 2 tool

The machine learning model will be trained on both public dataset and the captured images because in the dataset availableto us does not have specific and enough partial images to train the proposed model. A depth image example is shown in Fig.3 captured by the Structure Sensor camera.

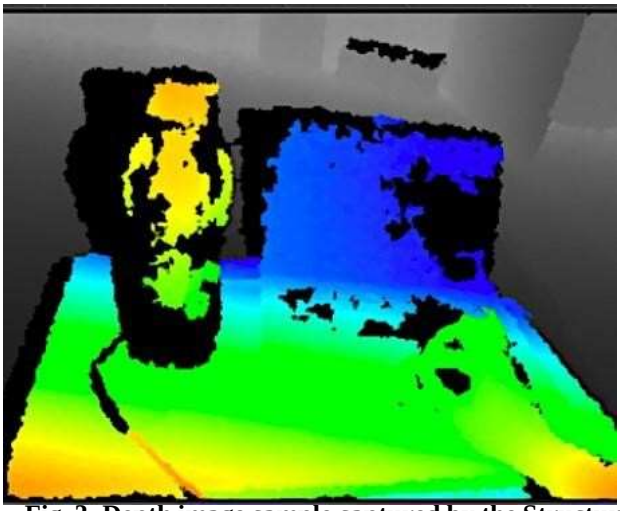

Fig. 3. Depth image sample captured by the Structure Sensor camera

1. OpenNI

OpenNI (Open Natural Interaction) is an open-source software. It provides access to Natural Interaction Devices. Natural Interaction Devices are the devices to get the image ofthe body or objects. The Structure Sensor camera is used to getthe 3D images of the object using OpenNI 2.

2. Structure Sensor/Core

It is a sensor to obtain the 3D model of and object and body. The Structure Sensor has a resolution of VGA 640x480or QVGA 320x240 with the precision of 0.5mm at 40cm and 30mm at 3m and with field of view horizontal 58-degrees and 45-degrees vertically. The Structure Core can produce better resolution images than the Structure Sensor camera, and it generates a resolution of 1,280x800 with depth image information at 60 FPS.

B. Pre-processing

For object detection, pre-process the image may be neededto increase the accuracy of our model after image capturing, because whenever image is captured by 3D camera there will be noise in the image. For the proposed model, the points of interests of environmental challenges are in cluttered, occluded, and images in low contrast light. To recognize and detect the 3D-object in the given challenges and to get their features preprocess may be required. Although, pre-process may increase the computation time and complexity of the research work, but without this part success rate may not be achieved. In the

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 8| Issue: 7| July 2022|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2022: 8.205 || ISI Value: 1.188

proposed method, OpenCV or Scikit Image will be used. DCNN can also be used for feature extraction, the first priority for feature extraction method would be DCNN. As of now, it cannot be concluded to one method, as all are renowned ways to perform pre-processing and feature extraction. These methods are compatible with Python and both of them supports machine learning libraries and deep learning frameworks. To obtain the best results, it will be tested with and without preprocessing the proposed datasets.

1. OpenCV:

OpenCV (Open source computer vision) is an open-source library for cross-platform. It is well known for computer vision, mainly for real-time functioning. It includes a machine learning library and deep learning framework like TensorFlow. Its applications are object identification, face recognition, gesture recognition, segmentation, feature extraction and many more.

2. Scikit Image:

It is an open-source Image Processing library for Python programming, specially designed for feature extraction, segmentation, analysis, filtering and color space manipulation and much more. In object detection, it is preferred to use NumPy, TensorFlow, and Scikit Learn. It is compatible with all of these libraries which makes it efficient and computationally fast.

C. Machine Learning Model and Classification

In the machine learning model, Convolutional Neural Network will be developed using Python, the CNN model is the best suited for image data classification. After the image has been pre-processed by using an appropriate tool, it will be served to the different layers of the DCNN model. In the CNN model, there are numbers of hidden layers which is made up of neurons. A simple block diagram of neural network for the proposed methods is shown in Fig. 4 and Fig. 5.

To classify the object image DCNN is the state-of-art technology, there are number of renowned DCNN architectures available to classify and detect the object. For the proposed design, some changes in existing architecture will be made to get better results for object detection and warehousing automation. As of now, a specific DCNN architecture is not fixed, as its output varies according to the requirements of an application and the image input. However, the proposed model will be tested with three different well known and competitive architecture which are VGGNet, Inception, and ResNet.

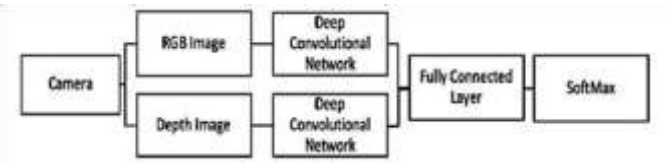

Fig. 4. Proposed method 1 process design block diagram

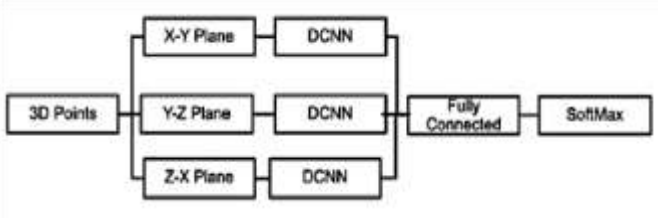

Fig. 5 Proposed method 2 process design block diagram

For the first proposed method, it is desired to experiment with two different DCNN models to obtain the best accuracy in detecting objects. After capturing the RGB-D images from the camera and pre-processing, each image will be fed to the twoseparate parallel DCNN architecture for convolution and feature extraction. Before going to the next layer, it is required to merge the output of both DCNN. Afterwards, it will beconnected to fully-connected layers, which will be used for final classification.

In the second proposed method, 3D point clouds will be divided into three different 2D-planes and it will be fed to the three different parallel 2D-DCNN architecture. Three different2D-Deep Convolutional Neural Network are used for one 3Dimage. After this step, output of the three DCNN models areconcatenated to a fully-connected and SoftMax layer to classify and recognize the object. After experimenting with these two methods, the output should be converted to the coordinate (*x, y, z*) form, which will be fed to the robot arm for further robot path planning and grasping algorithm to grasp theobject.

To train the deep layers for the DCNN architecture with large image datasets, it will require significant amount of computational time in where it can be reduced by using GPUs (Graphic processing unit) and the TensorFlow framework. TensorFlow is an open-source library for numerical computation that makes deep learning optimized and faster in computation. Furthermore, the proposed model can also be optimized using the Tensor Inference model to increase the computation efficiency using GPU's. Finally, the recognized and detected object position will be fed to the robot armcontroller which will be in coordinate (*x, y, z*) form. The givencoordinate values will help the robot arm to grasp the object from correct position using gripper.

D. Experimental procedures

In both of the proposed method, it is mentioned that separate parallel DCNN architecture will be used. There are many ways to do experimentation with the proposed model, either deep convolutional layers can be designed separately for both of the paths one for RGB and depth or similar layers can be used for both of them, similarly for the second proposed method. Experimentation also includes image pre-processing, as described in pre-processing section it will be part of experiments because accuracy changes with or without pre- processing. It is observed in literature review that sometimes a model receives better accuracy without preprocessing of image as compared to pre-processed image, because of scaling, converting normal images to edge, or contour form may lose some importantfeatures that loses overall accuracy of the model. To make sure the obtained

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 8| Issue: 7| July 2022|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2022: 8.205 || ISI Value: 1.188

results are accurate enough for the proposed work, experimentation will be done with pre-processing.

For method one, the RGB and depth will be used, and both are different form of images, so it is preferred to design separate architecture for both the paths to achieve the desired accuracy. For the second method mentioned, if this method is to be opted for the research work then experimental approach would be different as this method includes three different plane images, but the type of images will be in same form, therefore a single type of deep neural architecture can be selected. After processing the images and designing the DCNN architecture, all parameters will be optimized after experimenting with different learning rate and number of epochs for curve fittings.

CONCLUSION AND FUTURE WORK

Deep learning and CNN for object recognition and detection are reaching a new fields, and when it is combined with a robot arm, it can bring productive changes to the industries as working manpower is decreasing and robot can take the place of human. In this proposed research work, the deep neural network model is proposed for 3D object recognition and detection using RGBD images and in the point cloud form for pick and place robot in different challenging conditions. It is expected that with this DCNN proposed methods state of the art results could be achieved with the advantage of 2D computational powerful methods with 3D information.

The main objective of the proposed project is to detect the object using robotic vision under the challenging conditions like illumination, partial, occluded images, and send the controlling commands to the robot arm. The 3D-object image will be the captured by the Structure Sensor/Core camera mounted on robot arm. Different methods are discussed which is planned to implement in the proposed design, first one is to divide RGB-D image into two separate image RGB and Depth image which will be fed to separate parallel deep neural network, after these process both network will be combined to obtain desired output. In another method, 3D point cloud can be divided into three different planes and will be fed to the three different parallel deep neural networks for processing, then output of these three architectures will be merged and connected to fully-connected layer to obtain the output. The output of both the methods needed in coordinate (x, y, z) form to give the grasping command to the robot arm.

For future work, robot and machine learning is going to transform the way humans work impacting industries with more with robots than engineers. Just to make a robot reliable and harmless, more accurate vision is needed with better accuracy. Furthermore, this model can be implemented to the industrial robot after rigorous testing in an industrial environment, which can get higher accuracy for partial images and more objects can be added to the quality training with larger dataset. For the current grasping detection, separate grasping planning is required which can be improved, and combined with the same model. Other technique can also be implemented for training robot using reinforcement learning to make it more robust.

REFERENCES

[1]. *David Griffiths and Jan Boehm, "A Review on Deep Learning Techniques for 3D-Sensed Data Classification," Remote sense (2019) 11, 1499 www.mdpi.com/journal/remotesensing, [Accessed: Oct. 20, 2019]*

[2]. Sulabh Kumara and Christopher Kanan, "Robotic Grasp Detection using Deep Convolutional Neural Networks," 2016 International Conference on Intelligent Robots and Systems (IROS), Sept. 2017, Vancouver, BC, Canada

[3]. A. Garcia-Garcia, "PointNet: A 3D-Convolutional Neural Network for Real-Time Object Class Recognition," 2016 International JointConference on Neural Networks (IJCNN), Nov 2016, Vancouver, BC, Canada

[4]. Wikipedia, www.wikipedia.org [Accessed: Oct. 28,2019]

[5]. Nikolaus Correll, "Analysis and Observations from the First Amazon Picking Challenge," IEEE Transactions on Automation Science and Engineering, Volume: 15, Issue: 1, Jan. 2018

[6]. Shehan Caldera, Alexander Rassau and Douglas Chai, "Review of Deep Learning Methods in Robotic Grasp Detection," Multimodal Technologies and Interact. (2018) 2, 57, www.mdpi.com/journal/remotesensing, [Accessed: Oct. 20, 2019]

[7]. Yulan Guo, "3D-Object Recognition in Cluttered Scenes with Local Surface Features: A Survey," IEEE Transactions on Pattern Analysis and Machine Intelligence (Volume: 36, Issue: 11, Nov. 1, 2014)

[8]. Dogancan Temel, Jinsol Lee, "CURE-OR: Challenging Unreal and Real Environments for Object Recognition," 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Jan 2019, Orlando, FL, USA

[9]. SUN3D-database, http://sun3d.cs.princeton.edu [Accessed: Oct. 27,2019] SUN RGB-D: An RGB-D Scene Understanding Benchmark Suite

[10]. SUN RGB-D: An RGB-D Scene Understanding Benchmark Suite, http://rgbd.cs.princeton.edu, [Accessed: on Oct. 27, 2019]

[11]. RGB-D Object Dataset, https://rgbd-dataset.cs.washington.edu [Accessed: Oct. 27, 2019]

[12]. F. Gomez-Donoso, "LonchaNet: A Sliced-based CNN Architecture for Real-time 3D-Object Recognition," 2017 International Joint Conference on Neural Networks (IJCNN), May 2017, Anchorage, AK, USA

[13]. W. Czajewski, K. Kołomyjec, "3D-Object Detection and Recognition for Robotic Grasping Based on RGB-D Images and Global," Foundations ofComputing and Decision Sciences, Vol. 42, 2017

[14]. Daniel Maturana and Sebastian Scherer, "VoxNet: A 3D-Convolutional Neural Network for Real-Time Object Recognition," 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) Congress Center Hamburg, Sept 2015, Hamburg, Germany

[15]. Jean Lahoud, Bernard Ghanem, "2D-Driven 3D-Object Detection

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 8| Issue: 7| July 2022|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2022: 8.205 || ISI Value: 1.188

in RGBD Images," 2017 IEEE International Conference on Computer Vision (ICCV), Oct. 2017, Venice, Italy

[16]. Jonschkowski, "Probabilistic Multi-Class Segmentation for the Amazon Picking Challenge," 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Oct. 2016, Daejeon, South Korea

[17]. Carlos Herna´ndez Corbato, "Integrating Different Levels of Automation: Lessons from Winning the Amazon Robotics Challenge 2016," IEEE Transactions on Industrial Informatics, Volume: 14, Issue: 11, Nov. 2018

[18]. Ester Martinez-Martin, Angel P. del Pobil, "Object Detection and Recognition for Assistive Robots," IEEE Robotics & Automation Magazine, Volume: 24, Issue: 3, Sept. 2017

[19]. Xi Chen, Jan Guhl, "Industrial Robot Control with Object Recognition based on Deep Learning," 7th CIRP Conference on Assembly Technologies and Systems, www.sciencedirect.com, 2018, Stuttgart, Germany

[20]. Max Schwarz, "NimbRo Picking: Versatile Part Handling for Warehouse Automation," 2017 IEEE International Conference on Robotics and Automation (ICRA), May 2017, Singapore, Singapore

[21]. Colin Rennie, "A Dataset for Improved RGBD-Based Object Detection and Pose Estimation for Warehouse Pick-and-Place," IEEE Robotics and Automation Letters (Volume: 1, Issue: 2, July 2016)