# A Review of Deep Convolutional Neural Networks in Robotic Grasps Detection

Shehan Caldera
School of Engineering
Edith Cowan University
Australia
shelessa@our.ecu.edu.au

A/Prof. Alex Rassau
School of Engineering
Edith Cowan University
Australia
a.rassau@ecu.edu.au

Dr Douglas Chai
School of Engineering
Edith Cowan University
Australia
d.chai@ecu.edu.au

## ABSTRACT

The scientific community has recently experienced numerous developments in the area of intelligent machines because of the rapid improvement in data-based learning algorithms. Presently, Deep Learning (DL) is identified as capable of classification, recognition, and localisation applications of computer vision. Many recent studies have focused on implementing learning algorithms in robotic applications in Unstructured Environments (UE). Due to the variable nature of a UE, analytical robotic solutions can be expensive. Deep Convolutional Neural Network (DCNN) is the generic term for most of these learning algorithms. This paper reviews DCNN algorithms that appear in most recent robotic grasping work.

## 1. INTRODUCTION

Humans perform better than robots in learning to grasp objects.

## 2. GRASP REPRESENTATION

General purpose robots in a human-robot collaborative work environment need to have the ability to manipulate objects in the physical world. Object grasping is challenging in its own factors such as different object shapes and unlimited object poses. Successful grasp detection system should be able to overcome this challenge to predict valuable results. Unlike robots, humans almost immediately know how to grasp a given novel object. Robotic grasping performs well below the human object grasping benchmarks but continuous effort is made to improve considering the higher demand. Robotic grasping has 3 stages [1].

1. Grasp detection
2. Grasp planning
3. Hardware Control

Grasp detection recognises the grasping point(s) of a given image including the pose of the grasp in the image [2]. The pose describes the grasp orientation on the image plane. Grasp planning discusses the path planning that is required to achieve a successful grasp including obstacle avoidance if necessary. Hardware control implies the necessary programming or interfacing required to achieve the robotic manipulation for object grasping. As we humans use vision in such tasks as object grasping, robots use visual sensors or perception in the same task. Therefore grasp detection can be defined as given a dataset from a perception sensor, finding of the grasping points using ground truth data as a guideline if available. Different literature define robotic grasps for objects in various manners. Some of the promising approaches will be reviewed here.

Using a supervised learning approach, Saxena et al. [3] investigated a regression learning method to infer the 3-d location of the grasping point in a Cartesian coordinate system. Considering the camera position uncertainty into account Saxena et al. [3] also used a probabilistic model over possible grasping points. To further in their investigation, they had descritised the the 3-d workspace in order to find the grasping point $g$, given by

$$g = (x, y, z) \tag{1}$$

To infer the 3 coordinates, they used two or more images of the same object [3]. As shown in Figure 1, furthering their investigation, Saxena et al. [4] defined graspable regions of objects as grasping points for their learning algorithm to infer a 3-d grasping point. Most of these approaches were

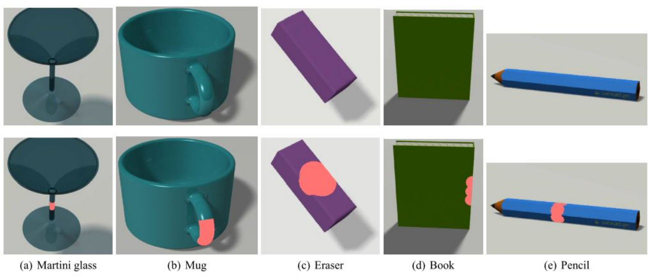


**Figure 1: Five object classes used for training with their grasping points marked on the images. Objects are a Martini glass, Mug, Eraser, Book, and a Pencil [4].**

investigated with RGB colour images and RGB simulated images before the introduction of the depth sensors which has improved all detection work including the grasp detection since then.

Considering the challenge of grasping novel objects, one first has to define the problem space accurately to tackle the problem in a proper manner. Point defined grasps only suggest where to grasp it. It does not determine how wide the gripper ends have to be opened nor the orientation of the gripper ends. Taking this into account Jiang et al. [5] has proposed a method to detect robotic grasp configurations in 3D space using RGB-D images. According to Jiang et al. [5] a grasping configuration has a seven dimensional representation. The grasp representation from Jiang et al. [5] contains **Grasping point**, **Grasping orientation**, and **Gripper opening width**. In 3D space the actual grasp representation, $G$ can be stated as in Equation 2.

$$G = (x, y, z, \alpha, \beta, \gamma, l) \tag{2}$$

Jiang et al. [5] represented a grasp as an oriented rectangle in the image plane. The rectangle edges shown in Figure 2 in blue colour represented the jaws of the gripper. The red coloured edges represented the opening or closing width of the gripper along with the direction of the motion. Simplifying the seven dimensional grasp rectangle representation by Jiang et al. [5], Lenz et al. [6] proposed a five dimensional representation similar to the approach by Jiang et al. [5]. This was based on the assumption of a good 2D grasp being able to be projected back to 3D space. While Lenz et al. [6] failed to evaluate their approach, Redmon et al. [7] confirmed the validity of the method with their own results. Redmon et al. [7] reassured the statements by Jiang et al. [5] and Lenz et al. [6], that detection of grasping points in this manner was analogous to object detection methods in computer vision but with an added term for the gripper orientation.

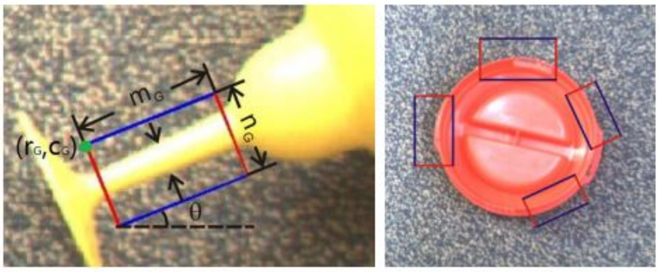


**Figure 2: Grasping rectangle representation. The upper-left corner $(r_G, c_G)$, length $m_G$, width $n_G$ and its angle from the x-axis, $\theta_G$. For some objects, like a red coloured lid in the right image, there can be multiple possible grasping rectangles [5].**

Adapting the method of [5, 6], Redmon et al. [7] presented a slightly updated representation of a grasp rectangle as shown in Figure 3. This modified rectangle grasp representation has been used few late publications proving its usefulness. In their work to use deep learning algorithms for robotic grasping detection, Sulabh et al. [1] have used the grasp rectangle proposed by Redmon et al. [7]. A very recent online project page [7] has cited the Redmon grasp rectangle in their work.

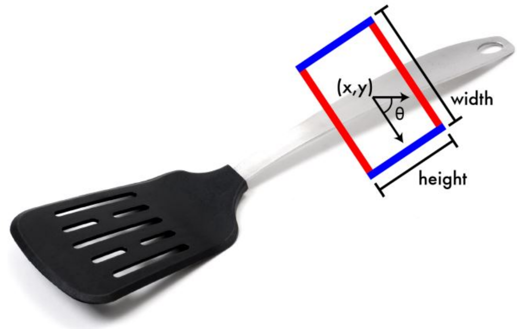


**Figure 3: Grasp rectangle by Redmon et al. [7]. Where $(x, y)$ is the center of the rectangle, $\theta$ is the orientation of the rectangle relative to the horizontal axis, $h$ is the height, and $w$ is the width [7].**

Inspired by the methods of Reinforcement Learning [**?**] Pinto et al. [**?**] presented a self-supervising algorithm for collecting data for learning to detect robotic grasps. Pinto et al. [**?**] stated that manually labelled data would not be scalable for various applications therefore application specific data is need for the individual application. Considering the amount of data that might require for such a method, Pinto et al. [**?**] presented a minimised representation of the grasp neglecting the gripper opening width. Given the grasping point $G = (x, y)$, Pinto et al. [**?**] prposed a method to predict the grasp orientation $\theta$. According to Pinto et al. [**?**] a grasp, $G$ can defined as $G = (x, y, \theta)$.

## 3. GRASP DETECTION

## 4. NEURAL NETWORK ARCHITECTURES

## 5. ROBOTIC GRASPING DATABASES

## 6. GRASP PLANNING AND SIMULATION

Also include results of the published work

## 7. CONCLUSION

## 8. REFERENCES

[1] S. Kumra and C. Kanan, “Robotic grasp detection using deep convolutional neural networks,” *CoRR*, vol. abs/1611.08036, 2016.

[2] 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,” *CoRR*, vol. abs/1703.09312, 2017.

[3] A. Saxena, J. Driemeyer, J. Kearns, and A. Y. Ng, “Robotic grasping of novel objects,” in *Advances in Neural Information Processing Systems 19* (B. Schölkopf, J. C. Platt, and T. Hoffman, eds.), pp. 1209–1216, MIT Press, 2007.

[4] 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.

[5] Y. Jiang, S. Moseson, and A. Saxena, "Efficient grasping from rgbd images: Learning using a new rectangle representation," in *2011 IEEE International Conference on Robotics and Automation*, pp. 3304–3311, May 2011.

[6] I. Lenz, H. Lee, and A. Saxena, "Deep learning for detecting robotic grasps," *Int. J. Rob. Res.*, vol. 34, pp. 705–724, Apr. 2015.

[7] J. Redmon and A. Angelova, "Real-time grasp detection using convolutional neural networks," *CoRR*, vol. abs/1412.3128, 2014.

[8] M. Basalla, F. Ebert, R. Tebme, and W. Ke, "Grasping for the real world (Greifen mit Deep Learning)." https://www.frederikebert.de/abgeschlossene-projekte/greifen-mit-deep-learning. Accessed: 2018-02-01.