Real-Time Robotic Grasping and Localization Using Deep Learning-Based Object Detection Technique

1st Mohannad Farag
Faculty of Engineering Technology
Universiti Malaysia Pahang
26300 Gambang, Pahang, Malaysia
mohanadfaraj@gmail.com

2nd Abdul Nasir Abd Ghafar
Faculty of Engineering Technology
Universiti Malaysia Pahang
26300 Gambang, Pahang, Malaysia
abdnasir@ump.edu.my

3rd Mohammed Hayyan ALSIBAI
Faculty of Engineering & Technology
International University for Science and Technology
Damascus , Syria
mhdhayyan@gmail.com

***Abstract*—This work aims to increase the impact of computer vision on robotic positioning and grasping in industrial assembly lines. Real-time object detection and localization problem is addressed for robotic grasp-and-place operation using Selective Compliant Assembly Robot Arm (SCARA). The movement of SCARA robot is guided by deep learning-based object detection for grasp task and edge detection-based position measurement for place task. Deep Convolutional Neural Network (CNN) model, called KSSnet, is developed for object detection based on CNN Alexnet using transfer learning approach. SCARA training dataset with 4000 images of two object categories associated with 20 different positions is created and labeled to train KSSnet model. The position of the detected object is included in prediction result at the output classification layer. This method achieved the state-of-the-art results at 100% precision of object detection, 100% accuracy for robotic positioning and 100% successful real-time robotic grasping within 0.38 seconds as detection time. A combination of Zerocross and Canny edge detectors is implemented on a circular object to simplify the place task. For accurate position measurement, the distortion of camera lens is removed using camera calibration technique where the measured position represents the desired location to place the grasped object. The result showed that the robot successfully moved to the measured position with positioning Root Mean Square Error (0.361, 0.184) mm and 100% for successful place detection.**

Keywords*—deep learning, CNN, object detection, edge detection, real-time grasp detection, robot positioning, robot arm

I. Introduction

Selective Compliant Assembly Robot Arm (SCARA) robot was developed as a new concept for assembly robots by Prof. Hiroshi Makino at University of Yamanashi in 1981 [1]. SCARA robot could move its gripper to any position within a circular space defined as its work envelope. As it is built based on a serial architecture, which means that the first motor carries the other motors, it requires a small footprint. On the other hand, SCARA controlling software requires inverse kinematics for linear movement which may increase its design cost [2]. The performance of SCARA robots, which are most adept in pick-and-place tasks, could be improved by increasing their speed, precision and capability [3]. Despite huge researches on computer vision, yet their impact on the robotic applications in industry is not very significant [4].

Recently, deep Convolutional Neural Networks (CNNs) have effectively affected deep learning models for different image processing applications such as object detection within input images [5], [6]. Transfer learning is an effective and fast way to build a CNN model rather than design and train it from scratch. In this approach, a pre-earned knowledge is adjusted and implemented to perform a new desired task such as robot grasp [7]. This approach is important since most successful deep learning-based robotic grasping research has used transfer learning to achieve state-of-the-art results and AlexNet [8] has been widely implemented in these studies [9]. In addition, the features that have been earned over large-scale datasets are generic in nature and can be used for new deep learning-based applications [10], [11].

In this paper, a deep CNN model is trained for real-time object and grasp detection based on AlexNet CNN. Four thousand images are acquired, labeled and processed as a training dataset for the derided robot grasp application over 40 classes. In these classes two object categories are associated with 20 different positions. To place the grasped object, a combination of Zerocross and Canny edge detectors [12] has been used for position measurement of a circular object in undistorted images based on camera calibration process [13], [14]. For grasp-and-place experimental implementation we used GLOBOT KSS-1500 SCARA robot [15] equipped with FLIR Point Grey Chameleon3 camera which is supported by USB3 Vision toolbox in MATLAB [16].

II. Deep Learning-Based Object Detection for Robotic Grasping

A. Convolutional Neural Network Object Detection Model

The structure of CNN object detection model implemented in this research is called KSSnet and it is based on AlexNet pre-trained model [8]. The pre-trained model is tailored for robot grasping task by changing the dimensions of the final layers to match 40 lasses using Deep Network Designer MATLAB application [17]. Then, CNN model is fine-tuned on the new image dataset implementing transfer learning process [7].

B. Training Dataset

SCARA robots usually deal with regular shaped objects in industrial assembly lines. Two cylindrical-shaped objects (labeled as AO and BO) with different dimensions and color are selected to represent assembly objects in assembly lines. Table I shows properties of the selected objects. SCARA training dataset is created based on AO and BO object categories as follows. Twenty positions in world coordinate are defined and associated with each object category. Each pair of object category and associated position is considered as a class in training dataset.

ACKNOWLEDGMENT

The authors would like to thank University Malaysia PAHANG to support this work under post graduate research grant PGRS190359.

REFERENCES

[1] L. Westerlund, *The Extended Arm of Man: A History of Industrial Robot*. Informationsförlaget, 2000.

[2] D. McMorran, D. C. K. Chung, J. Li, M. Muradoglu, O. W. Liew, and T. W. Ng, "Adapting a Low-Cost Selective Compliant Articulated Robotic Arm for Spillage Avoidance," *J. Lab. Autom.*, vol. 21, no. 6, pp. 799–805, Dec. 2016.

[3] R. Alcala, Z. G. Arceo, J. N. Baterisna, J. O. Morada, and J. Oliver D. Ramirez, "Selective Compliance Articulated Robotic Arm (SCARA): Application of Inverse Kinematics on the Control of Pick and Place Using Microsoft Kinect Xbox 360," *Res. Gate*, 2018.

[4] D. Hossain and G. Capi, "Object Recognition and Robot Grasping : A Deep Learning based Approach," *Robot. Soc. Japan*, no. October, 2016.

[5] Y. LeCun, B. Yoshua, and H. Geoffrey, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[6] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects.," *Science (New York, N.Y.)*, vol. 349, no. 6245. pp. 255–60, 2015.

[7] S. Khan, H. Rahmani, S. A. A. Shah, and M. Bennamoun, *A Guide to Convolutional Neural Networks for Computer Vision*, 1st ed., vol. 8, no. 1. Morgan & Claypool, 2018.

[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Adv. Neural Inf. Process. Syst.*, pp. 1–9, 2012.

[9] S. Caldera, A. Rassau, and D. Chai, "Review of Deep Learning Methods in Robotic Grasp Detection," *Multimodal Technol. Interact.*, vol. 2, no. 3, p. 57, 2018.

[10] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "CNN features off-the-shelf: An astounding baseline for recognition," *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, pp. 512–519, 2014.

[11] H. Azizpour, A. S. Razavian, J. Sullivan, A. Maki, and S. Carlsson, "Factors of Transferability for a Generic ConvNet Representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 9, pp. 1790–1802, 2016.

[12] R. M. Yousaf, H. A. Habib, H. Dawood, and S. Shafiq, "A Comparative Study of Various Edge Detection Methods," in *2018 14th International Conference on Computational Intelligence and Security (CIS)*, 2018, pp. 96–99.

[13] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, 2000.

[14] S. Urban, J. Leitloff, and S. Hinz, "Improved wide-angle, fisheye and omnidirectional camera calibration," *ISPRS J. Photogramm. Remote Sens.*, vol. 108, pp. 72–79, Oct. 2015.

[15] "DMBH Co., Ltd," 2018. [Online]. Available: http://www.dmbh.co.kr/?ckattempt=1.

[16] MathWorks, "Point Grey Camera Support from Image Acquisition Toolbox," 2019. [Online]. Available: https://www.mathworks.com/hardware-support/point-grey-camera.html.

[17] MathWorks, "Deep Network Designer," 2019. [Online]. Available: https://www.mathworks.com/help/deeplearning/ref/deepnetworkdesigner-app.html?searchHighlight=Deep. [Accessed: 04-Feb-2019].

[18] T. Borkar and L. Karam, "DeepCorrect: Correcting DNN models against Image Distortions," *arXiv Prepr.*, May 2017.

[19] MathWorks, "Set Up Parameters and Train Convolutional Neural Network," 2019. [Online]. Available: https://www.mathworks.com/help/deeplearning/ug/setting-up-parameters-and-training-of-a-convnet.html. [Accessed: 07-Feb-2019].

[20] MathWorks, "Transfer Learning Using AlexNet," 2019. [Online]. Available: https://www.mathworks.com/help/deeplearning/examples/transfer-learning-using-alexnet.html. [Accessed: 07-Feb-2019].

[21] P. Kim, *MATLAB Deep Learning*. Berkeley, CA: Apress, 2017.

[22] K. P. Murphy, *Machine Learning a probabilistic perspective*. MIT Press, 2012.

[23] M. Farag, A. N. Abd Ghafar, and M. H. Alsibai, "Grasping and Positioning Tasks for Selective Compliant Articulated Robotic Arm using Object Detection and Localization : Preliminary Results," in *2019 6th International Conference on Electrical and Electronics Engineering (ICEEE 2019)*, 2019.

[24] D. L. Olson and D. Delen, *Advanced data mining techniques*, 1st ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008.

[25] C. Sammut and G. I. Webb, Eds., *Encyclopedia of Machine Learning*. Boston, MA: Springer US, 2010.

[26] X. Zhang, L. Wang, Y. Zhao, and Y. Su, "Graph-Based Place Recognition in Image Sequences with CNN Features," *J. Intell. Robot. Syst.*, pp. 1–15, Aug. 2018.

[27] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, Jun. 2006.

[28] POWERS and David M.W., "Evaluation: from Precision, Recall and F-measure to ROC, Informedness, Markedness and Correlation," *J. Mach. Learn. Technol.*, vol. 2, no. 1, pp. 37–63, Dec. 2011.

[29] F.-J. Chu and P. A. Vela, "Deep Grasp: Detection and Localization of Grasps with Deep Neural Networks," *arXiv Prepr.*, no. 1, 2018.

[30] S. Kumra and C. Kanan, "Robotic grasp detection using deep convolutional neural networks," *IEEE Int. Conf. Intell. Robot. Syst.*, vol. 2017–Septe, pp. 769–776, 2017.

[31] J. Mahler *et al.*, "Dex-Net 2.0: Deep Learning to Plan Robust Grasps with Synthetic Point Clouds and Analytic Grasp Metrics," *arXiv Prepr.*, Mar. 2017.

[32] M. Ben-Ari, *Principles of concurrent and distributed programming*, 2nd ed. Addison-Wesley, 2006.

[33] J. Martin, *Programming real-time computer systems*. Englewood Cliffs: Prentice Hall, 1965.

[34] J. Watson, J. Hughes, and F. Iida, "Real-world, real-time robotic grasping with convolutional neural networks," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2017, vol. 10454 LNAI, pp. 617–626.

[35] D. Guo, T. Kong, F. Sun, and H. Liu, "Object discovery and grasp detection with a shared convolutional neural network," *Proc. - IEEE Int. Conf. Robot. Autom.*, vol. 2016–June, pp. 2038–2043, 2016.

[36] J. Sung, S. H. Jin, I. Lenz, and A. Saxena, "Robobarista: Learning to Manipulate Novel Objects via Deep Multimodal Embedding," *arXiv Prepr.*, Jan. 2016.

[37] L. Pinto and A. Gupta, "Supersizing self-supervision: Learning to grasp from 50K tries and 700 robot hours," *Proc. - IEEE Int. Conf. Robot. Autom.*, vol. 2016–June, pp. 3406–3413, 2016.

[38] S. Levine, P. Pastor, A. Krizhevsky, J. Ibarz, and D. Quillen, "Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection," *Int. J. Rob. Res.*, vol. 37, no. 4–5, pp. 421–436, Apr. 2018.

[39] Q. Lu, K. Chenna, B. Sundaralingam, and T. Hermans, "Planning Multi-Fingered Grasps as Probabilistic Inference in a Learned Deep Network," *arXiv Prepr.*, pp. 1–16, 2018.

[40] T. Suzuki and T. Oka, "Grasping of unknown objects on a planar surface using a single depth image," in *IEEE/ASME International Conference on Advanced Intelligent Mechatronics, AIM*, 2016, vol. 2016–Septe, pp. 572–577.

[41] J. Redmon and A. Angelova, "Real-time grasp detection using convolutional neural networks," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*, 2015, pp. 1316–1322.

[42] I. Lenz, H. Lee, and A. Saxena, "Deep learning for detecting robotic grasps," *Int. J. Rob. Res.*, vol. 34, no. 4–5, pp. 705–724, Apr. 2015.

[43] A. Jabalameli, N. Ettehadi, and A. Behal, "Edge-Based Recognition of Novel Objects for Robotic Grasping," *arXiv Prepr.*, Feb. 2018.

[44] G. Huang, H. Lin, and P. Chen, "Robotic Arm Grasping and Placing Using Edge Visual Detection System," *2011 8th Asian Control Conf.*, vol. 240, pp. 960–964, 2011.