

Grasp Point Estimation using Simulator-Generated Datasets Including Pose Information

Ryoga Maruno

Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Tomoya Shiba

Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Naoki Yamaguchi

Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Hakaru Tamukoh

Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

*Email: maruno.ryouga396@mail.kyutech.jp, shiba.tomoya627@mail.kyutech.jp,
yamaguchi.naoki892@mail.kyutech.jp, tamukoh@brain.kyutech.ac.jp*

Abstract

We develop a system that automatically generates training datasets for object recognition models using a simulator. In this study, we incorporate pose information into the dataset. Using this system, we develop a method for estimating grasp points for objects that are difficult for robots to grasp, selecting a toy airplane as the target object. Three specific points are assigned to the object: the front, center, and back. In the grasp point estimation process, the center point is designated as the grasp point. The robot's arm achieves an appropriate grasp by moving perpendicularly to the line connecting the front and back points toward this grasp point.

Keywords: Object recognition, Pose estimation, Dataset generation, Grasp point

1. Introduction

In recent years, object grasping by robots has emerged as a critical application area in assistive robotics, enabling robots to perform tasks such as supporting daily activities, handling irregularly shaped objects, and aiding individuals with physical limitations. However, grasping accuracy often decreases for objects with complex shapes or unpredictable orientations, presenting significant challenges for conventional methods. Recent advancements in simulation technology provide a promising solution by enabling the automated generation of diverse and realistic training datasets under controlled conditions. This approach not only addresses the limitations of real-world data collection but also facilitates the development of robust models capable of handling a wide variety of objects.

This study proposes a method to improve grasping accuracy for challenging objects by incorporating pose information into object detection models and enabling keypoint detection. Additionally, the training dataset is automatically generated using a simulator. Building on this foundation, the study aims to develop a highly accurate and adaptable object detection model through simulation-based data generation, ultimately enhancing robotic grasping performance in real-world scenarios.

2. Related Work

2.1. Object Recognition

Using the generated dataset, the model was trained with the YOLO-based skeletal estimation model YOLO11 (parameters: 2.9M) [1]. About 100k images were used as training data, and about 25k images were used as validation data. The training was conducted with a batch size of 16 and for 60 epochs. This process established a foundation for evaluating the performance of the model based on the generated dataset.

2.2. YCB Object

For the implementation of the proposed method, YCB objects [2] were used as the target objects for validation. YCB objects are a standardized dataset widely used in research on object grasping and manipulation tasks, consisting of objects with diverse shapes, materials, and sizes. Therefore, they are well-suited for evaluating the performance of algorithms related to object recognition and robotic manipulation. In this study, representative objects were selected from the YCB dataset to perform keypoint annotation and evaluate estimation accuracy using the proposed method. Examples of YCB objects are shown in Fig. 1.



Fig. 1 Examples of YCB objects

2.3. Object Keypoint Similarity

OKS (Object Keypoint Similarity) (Eq. (1)) was adopted as the metric to evaluate the accuracy of keypoint estimation. OKS is a standard metric for measuring the accuracy of estimated keypoints on target objects, calculated based on the distance between the estimated keypoints and the ground truth data. Since the score computation involves weighting that considers the object's scale and the importance of the keypoints, it is applicable to objects of various sizes and shapes. In this study, OKS was calculated for all keypoints as well as for individual keypoints to evaluate estimation accuracy in detail. Table 1 provides the definitions of the symbols used in Eq. (1).

$$OKS = \frac{\sum_i \exp(-d_i^2 / 2s^2 k_i^2) \delta(v_i > 0)}{\sum_i \delta(v_i > 0)} \quad (1)$$

Table 1 the definitions of the symbols of OKS

Symbol	Meaning
d_i	Euclidean distance between the estimated keypoint and the ground truth keypoint
s	Scale of the object
k_i	Importance constant for the keypoint
$\delta(v_i > 0)$	Indicator function that equals 1 if the visibility ($v_i > 0$)

3. Proposed Method

In this study, we generate training datasets required for object detection models by performing 3D scanning of target objects and randomly placing them within a simulated environment [3]. Tobin et al. proposed *Domain Randomization (DR)* [4], demonstrating that training object detection tasks using only synthetic images can result in models capable of functioning in real robotic systems.

Additionally, Jonathan et al. [5] applied this technique to object recognition tasks, achieving performance in vehicle detection comparable to models trained on real-world images. These findings suggest that DR is expected to minimize Domain Gap effectively. The Domain Gap in this study refers to the difference between simulation data and real-world image data.

Our dataset generation process adopts DR to create diverse data by randomly altering backgrounds, object positions, and camera angles. An example of the generated data is shown in Fig. 2. During this process, annotation data, such as object position information and labels is automatically stored.

Furthermore, in this study, we added a function to generate annotation data with keypoint information for each object to improve object grasping performance. In addition, we implemented a function to determine keypoint visibility by extending rays from the simulator's camera position to each keypoint and verifying their visibility by checking for potential obstructions, making the keypoint information more detailed and useful. Fig. 3 shows an example of the visualized keypoint information. In Fig. 3, the positions and visibility of the keypoints are clearly indicated, providing a concrete understanding of the characteristics of the training dataset.

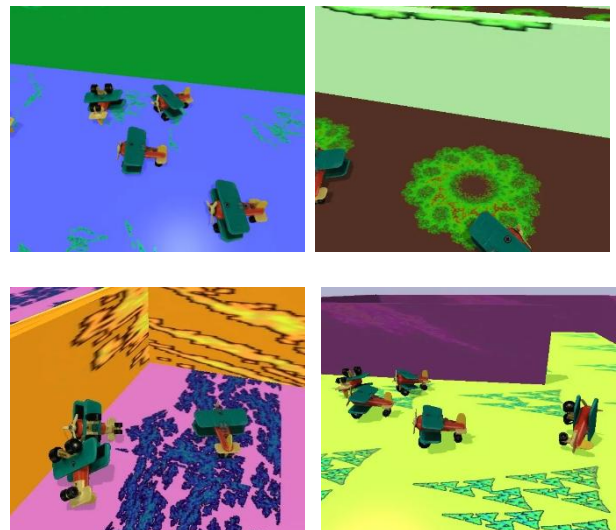


Fig. 2 The Training Datasets Generated in the Simulator

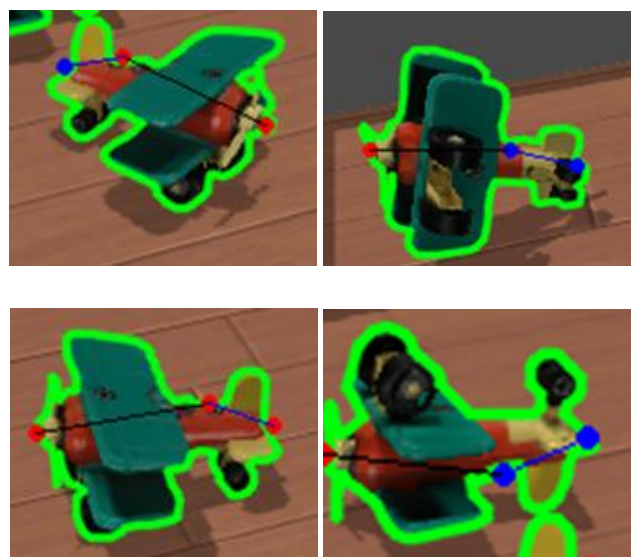


Fig. 3 Keypoint Annotations with Visibility in the Simulated Dataset (Red: Visible, Blue: Non-Visible)

4. Experiment

The purpose of this experiment is to verify the accuracy of keypoint estimation using the proposed method. The reason for selecting the toy-airplane as the target object is that it is one of the objects with low grasping success rates in conventional methods, such as Grasp Pose Detection (GPD) [6] and its improved version, PointNetGPD [7], making it suitable for evaluating the effectiveness of the proposed method. For validation, the toy-airplane was randomly placed in plausible configurations within a simulator, and 10,000 generated images were used. An example of the test data is shown in Fig. 4.

The toy-airplane is annotated with three keypoints: the front, center, and back. The center is defined as the grasping point, and the arm is assumed to be inserted perpendicularly to the vector formed by the front and back points. For evaluation, the mean OKS of all keypoints, the mean OKS of the keypoint defined as the grasping point, and the angular error between the vector formed by the front and back points and the annotated data were calculated. This evaluation design allows for a detailed verification of the effectiveness of the proposed method for objects that are difficult to grasp.

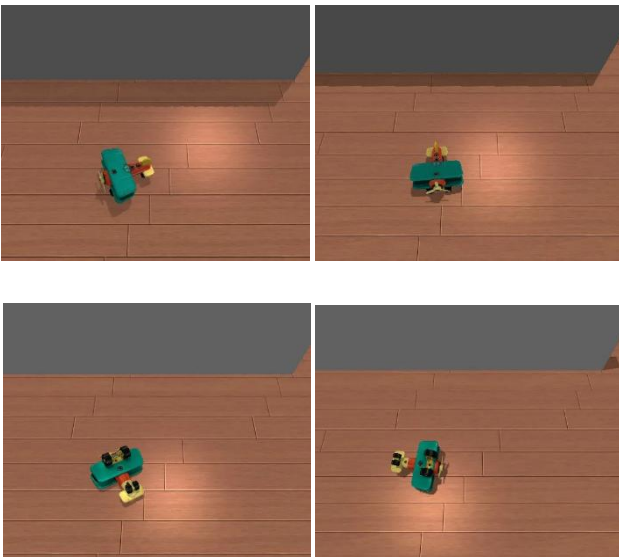


Fig. 4 The Test Datasets in the Simulator

5. Results

The results are shown in Table 2. The proposed method demonstrated excellent keypoint estimation performance for objects like the toy-airplane, which are difficult to grasp. In particular, the high OKS values and small angular errors suggest that keypoint information is effective for grasping motions. This indicates that the proposed method has the potential to optimize the insertion angle and position of the robotic arm even for objects that have been considered difficult to grasp using conventional methods. Fig. 5 shows visualized outputs of the grasping points and arm angles.

On the other hand, it is important to note that the validation data were generated within a simulator. The extent to which the proposed method can maintain its performance when faced with the diversity and uncertainty of objects in real-world environments requires further investigation. Additionally, verifying whether similar effects can be achieved for objects other than the toy-airplane is also essential.

Table 2 Keypoint Estimation Performance for Toy-Airplane Using the Proposed Method

	Result(average)
OKS(all keypoints)	0.9996
OKS(front)	0.9995
OKS(center)	0.9996
OKS(back)	0.9996
Angle error	0.0215

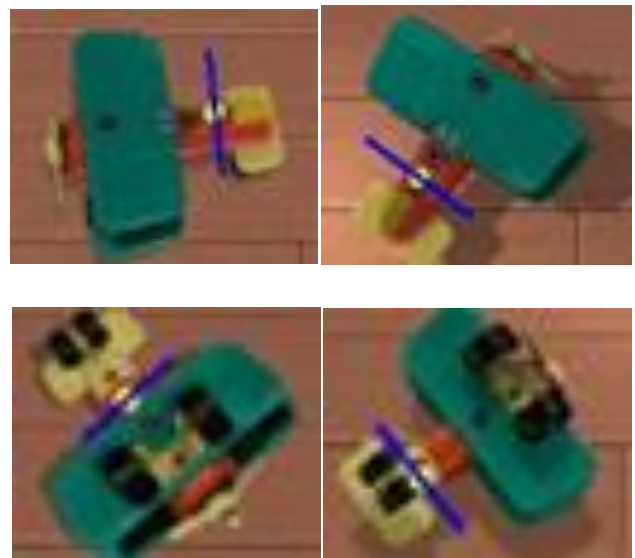


Fig. 5 Grasp Point and Insertion Angle Output

6. Conclusion

This study proposed a method to enable accurate keypoint estimation for objects that are difficult to grasp, such as the toy-airplane, by generating a training dataset with keypoint annotations using a simulator. Experimental results showed favorable outcomes in terms of the mean OKS of all keypoints, the mean OKS of the keypoint defined as the grasping point, and angular error, demonstrating that the proposed method can effectively provide the information necessary for object grasping motions. This study offers a new solution for objects that were challenging to handle using conventional methods, contributing to the realization of flexible object manipulation by robots.

As a future prospect, the first step is to validate the proposed method using real images to improve its accuracy. Additionally, it is necessary to implement this

method on actual robots and evaluate its effectiveness in real-world environments through comparisons with conventional methods. Additionally, while the current approach focuses on 2D keypoints, exploring the direct acquisition of 3D keypoints or their conversion from 2D to 3D is expected to open possibilities for more advanced object manipulation [8][9].

7. References

1. Ultralytics. "Ultralytics YOLO11" 2024. Available at: <https://docs.ultralytics.com/>. Accessed: 2024-11-12.
2. Berk Calli, Arjun Singh, James Bruce, Aaron Walsman, Kurt Konolige, Siddhartha S. Srinivasa, Pieter Abbeel, and Aaron M. Dollar. "YCB Benchmarking Project: Object Set, Data Set and Their Applications." *Journal of The Society of Instrument and Control Engineers*, 56(10):792–797, 2017.
3. Tomohiro Ono. "Establishing Fundamental Technologies for Home Service Robots: Data Generation for Sim2Real and Its Application to Pick-and-Place" (Kateiyo Sabisu Robotto no tame no Kiban Gijutsu no Kakuritsu: Sim2Real wo Jitsugen suru Data Seisei to Pick-and-Place he no Oyo). Doctoral Dissertation, 2023. (Japanese)
4. Josh Tobin, Rachel Fong, Alex Ray, Jonas Schneider, Wojciech Zaremba, and Pieter Abbeel. "Domain Randomization for Transferring Deep Neural Networks from Simulation to the Real World." In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 23–30, 2017.
5. Jonathan Tremblay, Aayush Prakash, David Acuna, Mark Brophy, Varun Jampani, Cem Anil, Thang To, Eric Cameracci, Shaad Boochoon, and Stan Birchfield. "Training Deep Networks with Synthetic Data: Bridging the Reality Gap by Domain Randomization." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 969-977, 2018.
6. Andreas Ten Pas, Marcus Gualtieri, Kate Saenko, and Robert Platt. "Grasp Pose Detection in Point Clouds." *The International Journal of Robotics Research (IJRR)*, 36.13-14(2017): 1455-1473.
7. Hongzhuo Liang, Xiaojian Ma, Shuang Li, Michael Görner, Song Tang, Bin Fang, Fuchun Sun, and Jianwei Zhang. "PointNetGPD: Detecting Grasp Configurations from Point Sets." In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 3629-3635, 2019.
8. Ching-Hang Chen and Deva Ramanan. "3D Human Pose Estimation = 2D Pose Estimation + Matching." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7035–7043, 2017.
9. Diogo C. Luvizon, David Picard, and Hedi Tabia. "2D/3D Pose Estimation and Action Recognition Using Multitask Deep Learning." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5137–5146, 2018.

Authors Introduction

Mr. Ryoga Maruno



He received the B.Eng. degree from the National Institute of Technology, Kurume College, Japan, in 2024. He is a master's degree student at the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology. His research interests include dataset creation and Visualization.

Mr. Tomoya Shiba



He received the B.Eng. degree from National Institute of Technology, Kagoshima College, Japan, in 2021. He received the M.Eng. from Kyushu Institute of Technology, Japan, in 2023. He is currently in a Ph.D. student in the graduate school of Life Science and Systems Engineering, Kyushu Institute of Technology. His research interest includes image processing, motion planning, and domestic service robots.

Mr. Naoki Yamaguchi



He received the B.Eng. degree from the National Institute of Technology, Ube College, Japan, in 2023. He is a master's degree student at the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology. His research interests include dataset creation and Visualization.

Dr. Hakaru Tamukoh



He received the B.Eng. degree from Miyazaki University, Japan, in 2001. He received the M.Eng and the Ph.D. degree from Kyushu Institute of Technology, Japan, in 2003 and 2006, respectively. He was a postdoctoral research fellow of 21st century center of excellent program at Kyushu Institute of Technology, from April 2006 to September 2007. He was an assistant professor of Tokyo University of Agriculture and Technology, from October 2007 to January 2013. He is currently an associate professor in the graduate school of Life Science and System Engineering, Kyushu Institute of Technology, Japan. His research interest includes hardware/software complex system, digital hardware design, neural networks, soft-computing and home service robots. He is a member of IEICE, SOFT, JNNS, IEEE, JSAI and RSJ.
