The Early Deployment of Robot Picking Systems Using Auto Picking Point Annotation

Cheng-Han (Lance) Tsai, Jen-Yuan (James) Chang, Eduin E. Hernandez, Luis F. Garcia

***Abstract*— Artificial Intelligence (AI) with computer vision has been widely used in robotic pick and place operations for detecting the coordinates of the workpieces. One of the biggest bottlenecks of AI is the need for a vast amount of labeled data for sufficient training of the AI model. If the quantity of the data is insufficient or if the quality of the labeling is unstable, problems will arise when training and testing the AI. To resolve this issue, we propose an early deployment method for automatically generating diverse data with domain randomization and an auto-picking point annotation system for labeling the data. This system will separate one workpiece into multiple parts and consider the suitable and non-suitable areas for picking which will be labeled with OK/NG respectively.**

***Keywords: Auto Annotation, object recognition, robotic random bin-picking*.**

I. Introduction

In comparison to AI with computer vision, the algorithm of a traditional computer vision is more likely to underperform. If there are any changes in the conditions such as lighting, texture, or scaling, then the desired rule-based solutions will be not reached. If a large enough dataset with diversity for the AI model's training is prepared, higher performance and robustness can be achieved by the model. YOLO [1], an object detection AI model, is extremely fast and accurate in detecting multiple and diverse objects as well as outputting their bounding box at their corresponding locations. MaskRCNN [2], an instance segmentation AI model, functions well in segmenting objects with pixel alignment providing a higher precision than that of a bounding box. The incorporation of AI models into vision has been widely used in robotic pick and place research [3] [4]. This incorporation facilitates the detection of an object without CAD models and promotes the rapid conversion for factory purposes since it does not require an engineer to rewrite the algorithm, it just requires a change in the datasets. Although the current results of using AI in the mentioned fields are decent, AI models require a vast amount of data and labeling of the corresponding answers for sufficient training. However, the labeling will require significant manpower. If the quality of the labels for the datasets is unstable, such as the occasional mislabeling, problems will arise when training and testing the AI model. To combat these issues, data augmentation has been used to increase image datasets [5] by creating slight alterations in existing data and Generative Adversarial Network (GAN) [6] has been used to automatically increase or generate new datasets based on the training dataset.

Ideally, it is better to generate the images through simulators [7] since they can generate the objects with random posture with a physics engine to calculate the physics reaction of objects in the environment. [8] proposes a method for mapping textures on an object with lower distortion. This optimization function provided is a new mapping approach in the computer graph domain. And [9] uses CoppeliaSim, a robot simulator platform, to simulate the motion of the robot along with the robotic picking, stacking, and truck unloading which is the three main jobs for robots in a warehouse.

The above literature contributes to reducing the need for manpower for AI. However, the methods still face the following problems [10]:

- Generation of the images without annotations cannot be used for the training of the AI model as it still requires the annotations for the learning phase.
- Annotations that do not consider the grasping point will fail at robotic picking.
- Data augmentation does not provide enough diversity.

II. System Architecture

A. System Structure

First, our system automatically generates the 3D model with the real workpiece's texture and loads the model into the simulator. Then, uses the simulator to generate data for a large quantity and diversity of images. Next, it calculates each object's occlusion status and provides a different annotation class accordingly (OK/NG). Lastly, the generated datasets and annotation files are used to train the AI model. The resultant AI model is then inserted and used by the robotic picking system as shown in Fig 1.

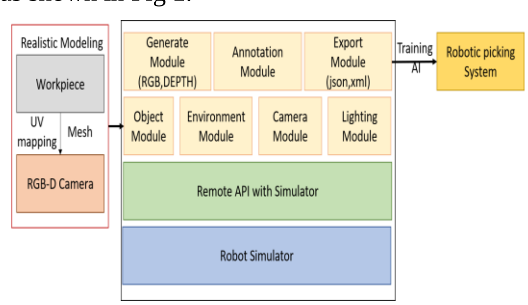

Figure 1. System Architecture

Research supported by the Industrial Technological Research Institute, TW.

C. H. Tsai is with the Department of Power Mechanical Engineering, National Tsing Hua University, TW (e-mail: LanceTSAI@itri.org.tw).

J. Y. Chang is with the Department of Power Mechanical Engineering, National Tsing Hua University, TW (e-mail: jychang@pme.nthu.edu.tw).

E. E. Hernandez is with the Electrical and Computer Science Department, National Yang Ming Chiao Tung University, TW (e-mail: eduin.ee08@nycu.edu.tw).

L. F. Garcia is with the Computer Science Department, National Yang Ming Chiao Tung University, TW (e-mail: luisgarcia.cs06@nycu.edu.tw).

B. Experimental Devices, Instruments, and Workpieces

The real case experiment test was performed using a 6 axis MITSUBISHI RV-13FLM-1D-S11 robot with Realsense D435 RGB-D camera and vacuum gripper with a 30mm diameter. The experimental workpieces used were combination wrenches, socket wrenches, and grocery items such as candy bags, chewing gum packages, and chocolate bars. The setup is shown in Fig. 2.

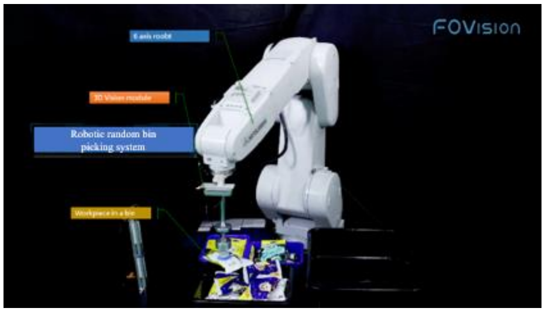

Figure2. Experimental Setup

III. PICKING MODELS

This section describes the picking model, which is generated using multiple RGB-D cameras. The workpiece is scanned to generate the 3D mesh file. A 2D image is extracted to map the texture on the 3D model using the UV-mapping method [8].

The mesh is made up of vertices, each having its X, Y, and Z coordinate, and a texture that is applied from an image. The mapping method used for the texture is UV-mapping, which finds the relationship between the 2D image pixel and the 3D coordinate of the mesh using the following geometric equations:

$$u = \sin\theta \cos\varnothing = \frac{X}{\sqrt{X^2+Y^2+Z^2}} \tag{1}$$

$$v = \sin\theta \sin\varnothing = \frac{Y}{\sqrt{X^2+Y^2+Z^2}} \tag{2}$$

Another approach is generating the models with only the picking surface of the real workpiece. To do this, we let the workpiece's picking surface face towards the camera and take the photo. We redo this process, but this time with a non-picking surface. Using this method, we can generate 3D models from a single real workpiece to multiple 3D models in a fast manner, including 3D picking models and 3D non-picking models. The modeling pipeline is shown in Fig. 3.

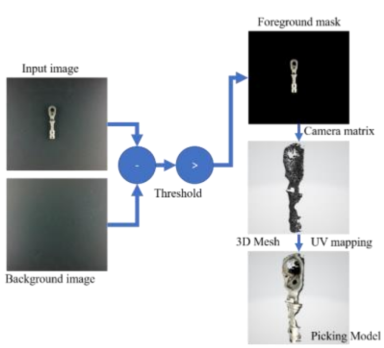

Figure 3. Picking Modeling Pipeline

A. Automatic Generation of Images

This section describes the automatic generation of the images and their corresponding annotations, using the picking models as described in Section III. A simulator with a physics engine and domain randomization with the following parameters are used: object posture, object quantity, lighting posture, camera postures, light intensity, light color, and color rendering.

After the physics reactions have settled in each environment, a shot is taken by the virtual camera inside the simulator and extracted to calculate each workpiece's occlusion status.

B. Priority Strategy for Labels

Like in the real case scenario where only the easily accessible workpieces are picked, the annotations are designed so the picking class for each workpiece considers the occlusion level and the visible pixel surface area seen by the camera as shown in Fig 4. For example, workpieces on the top of the crate will have a higher priority for picking because they have less occlusion and higher pixel surface area visible to the camera in comparison to the other workpieces, making the picking easier for the robot. Workpieces closer to the middle of the crate will have a priority higher than those which are closer to walls since these can be considered as obstacles, increasing the difficulty in the picking.

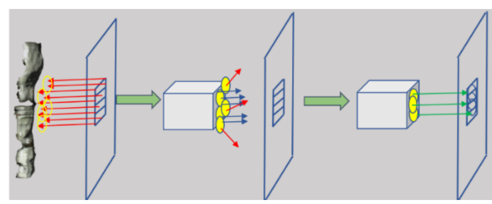

Figure 4. Finetuning the label based on the normal of the picking point

The smoothness of the picking point surface and the angle from which it is picked play a significant role in the success of picking the workpiece. A change in the elevation of the surface or insufficient surface area, as seen in Fig. 5, can prevent grippers from successfully picking up a workpiece.

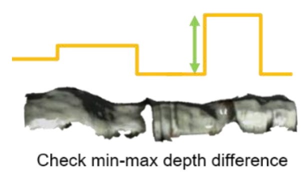

Figure 5. Finetuning the label based on the flatness of the picking point

Fig. 6 displays the schematic diagram from modeling different surfaces of the object. Several picking and non-picking surface models are made from a single object so the AI can learn and infer which workpiece can be picked and, when performing the picking, from what part can it be picked.

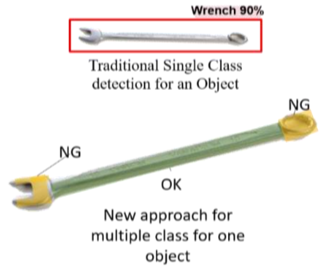

Figure 6. Multiclass Labeling for Single Workpiece

IV. Results

We use different workpieces for the real robot picking using the AI model trained with the datasets generated by our system. The experiment runs 500 picking cycles for the same type of workpieces to test the system's reliability. A picking cycle consists of selecting and picking a workpiece. It is considered successful if the robot picks up the selected workpiece within 3 attempts, as shown in Table I. The influence the success rate is the light reflect of workpiece like the wrench which material is made by metal with high lighting reflect, otherwise the average success rate are over 90% with the general workpiece.

TABLE I. Robot picking success rate

Workpiece	*Cycle*	*Success*	*Success rate*
Candy	500	440	88%
Chewing gum	500	420	84%
Chocolate	500	480	96%
Chess	500	490	98%
Wrench	500	400	80%

V. DISCUSSION AND LIMITATIONS

The advantage of this system is its ability to easily model the workpiece and automatically generate enormous amounts of labeled datasets within an hour. Training AI models with these data and annotations can speed up the learning process for robotic picking systems without the need to manually create or buy CAD models. It is factory flexible and supports rapid changeover for production. Unfortunately, the current simulator only supports rigid object models, thus simulating flexibility and elasticity when stacking objects is not supported by our systems like for workpieces such as line cables, candy bags, and chewing gum packages.

VI. Conclusion And Future Work

In this paper, we proposed a new approach for automating the AI pipeline which not only can be used to compete with other robotic picking systems but also serve as a tool to supply annotated datasets for AI research and robotic picking systems.

This paper proposes a method to generate the picking models with a realistic texture that considers its picking surface. It also automatically generates the images and annotations based on the occlusion status of workpieces for different class labels to teach the AI what are the best workpieces and picking points for the robot. The datasets generated by our system include a variety of real-world variables like the posture of the camera, light, and workpieces; color rendering; and light decay which considers environment changes to generate the datasets.

Lastly, we verified the performance of this system with the inference of the real robot for picking different workpieces which are completely modeled, generated, and annotated by our system.

Future works include expanding the picking annotation strategy by considering a wider variety of grippers like the jaw parallel grippers or hand-type grippers.

Acknowledgment

We would like to give thanks to Industrial Technology Research Institute (ITRI) for their support during this research.

References

[1] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Op-timal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934*, 2020.
[2] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961–2969.
[3] C.-T. Lee, C.-H. Tsai, and J.-Y. Chang, "A cad-free random bin picking system for fast changeover on multiple objects," in *Information Storage and Processing Systems*, vol. 83600. American Society of Mechanical Engineers, 2020, p. V001T02A002.
[4] F. Spenrath and A. Pott, "Using neural networks for heuristic grasp planning in random bin picking," in 2018 *IEEE 14th International Conference on Automation Science and Engineering (CASE)*. IEEE, 2018, pp. 258–263.
[5] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, pp. 1–48, 2019.
[6] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp.2223–2232.
[7] P. Jiang, Y. Ishihara, N. Sugiyama, J. Oaki, S. Tokura, A. Sugahara, and A. Ogawa, "Depth image–based deep learning of grasp planning for textureless planar-faced objects in vision-guided robotic bin-picking," *Sensors*, vol. 20, no. 3, p. 706, 2020.
[8] J. Maillot, H. Yahia, and A. Verroust, "Interactive texture mapping," in *Proceedings of the 20th annual conference on Computer graphics and interactive techniques*, 1993, pp. 27–34.
[9] F. Islam, A. Vemula, S.-K. Kim, A. Dornbush, O. Salzman, and M. Likhachev, "Planning, learning and reasoning framework for robot truck unloading," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 5011–5017.
[10] M. L. Gray and S. Suri, *Ghost work: How to stop Silicon Valley from building a new global underclass*. Eamon Dolan Books, 2019.