A New Approach to Enhance Artificial Intelligence for Robot Picking System using Realistic Digital Twins Tool

Cheng-Han (Lance) Tsai, Jen-Yuan (James) Chang, Eduin E. Hernandez, Xiu-Wen You

***Abstract*— Robotic bin picking system (RBP) with Artificial Intelligence (AI) has been widely used in different applications for learning the features of new workpieces and detecting their respective coordinates through a camera. 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 either the quantity of the data is insufficient or the quality of the labeling is unstable, problems will arise when training and testing the AI. Also, many of the robotic picking systems will use a vacuum as the main gripper because the soft vacuum cup can easily adapt to the workpiece, however because of the different speeds of the handling, weights of the workpieces, and the air pressure exerted by the vacuum cup there is diversity in the behavior of the grasping. Thus, it is challenging to find a digital twin system to verify and analyze RBP to improve its real-world performance.**

To resolve the first of the problems, 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. For the second problem, we employ the usage of mathematical formulas to calculate and approximate the gripper reactions using various parameters.

Our proposed system will train and implement an AI for RBP using its own generated dataset in an early stage. Furthermore, we will test its performance in the simulator and real-world with a vacuum gripper to validate our system and formulas.

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

I. Introduction

In comparison to AI with computer vision, the algorithm of 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 not be achieved. 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 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 some of these problems, 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. Authors in [8] propose 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 are the three main jobs for robots in a warehouse.

It is also critical to simulate RBP behavior when picking the workpiece to analyze its limitations. Authors in [11] propose mathematical formulas for unilateral grasping of vacuum cups, which can be used as theoretical limits for stable grasping. While in [12], authors propose a model for contact stability for vacuum cups during movement along with a procedure to calculate or approximate such model.

The above literature contributes to reducing the need for manpower for AI and predicting the RBP gripping stability. 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.

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).
X. W. YOU is with the electrical engineering Department, National Yang Ming Chiao Tung University, TW (e-mail: itri459217@itri.org.tw).

- Annotations that do not consider the grasping point will fail at robotic picking.
- Data augmentation does not provide enough diversity.
- Theoretical limits of the grasping stability are not tested and corrected.
- Other constraints for the stability of the grasping are not explored, i.e the angle of grasping or the diameter of the vacuum cup.

To combat this set of problems, we propose 1) an early deployment method for automatically generating diverse data with domain randomization and an auto-picking point annotation system for labeling the data, and 2) mathematical formulas to calculate and approximate the gripper reactions using various parameters.

This system will help outsource manpower in hazardous environments such as heat-treatment plants in Taiwanese factories, starting in 2022. In said scenario, the system will select and pick a heavy workpiece amongst others that are randomly stacked inside the bin. After picking, the industrial robot will move the workpiece towards the conveyor that will then carry it into the furnace. Because the workpieces are heavy, the workpiece types are numerous, and the environment surrounding the furnace is hot (~50°C), there is a need for rapid change over systems that can train and implement AI models from self-generated datasets and digital twins to analyze and predict the grippers reactions in the real world using simulators.

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.

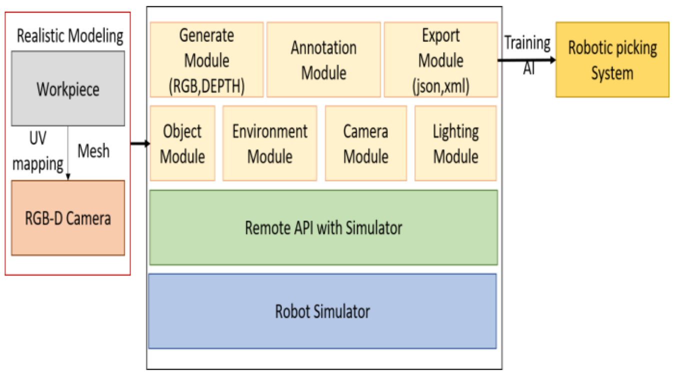

Fig. 1. System Architecture.

B. Experimental Devices, Instruments, and Workpieces

The real case experiment test was performed using a 6 axis M-710iC/45M robot with FOVision vision module shown in Fig. 2 and 3. The experimental workpieces used were candy bags, chewing gum packages, chocolate bars, checker pieces, and weight blocks.

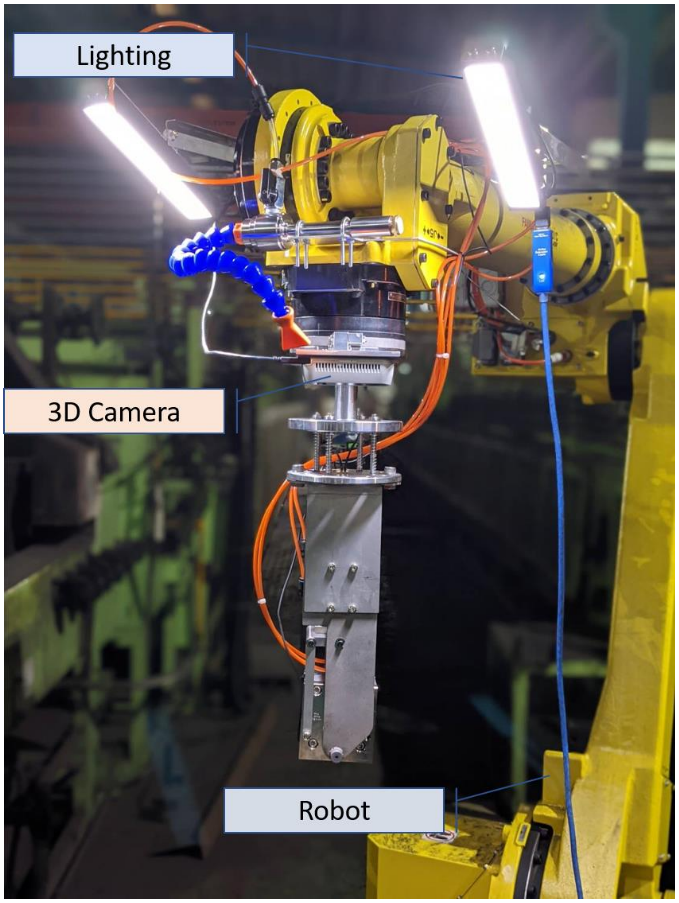

Fig. 2. M-710iC/45M Robot

Fig. 3. FOVision vision module.

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:

$$u = sin\theta cos\phi = \frac{X}{\sqrt{X^2 + Y^2 + Z^2}} \quad (1)$$

$$v = sin\theta sin\phi = \frac{Y}{\sqrt{X^2 + Y^2 + Z^2}} \quad (2)$$

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. 4.

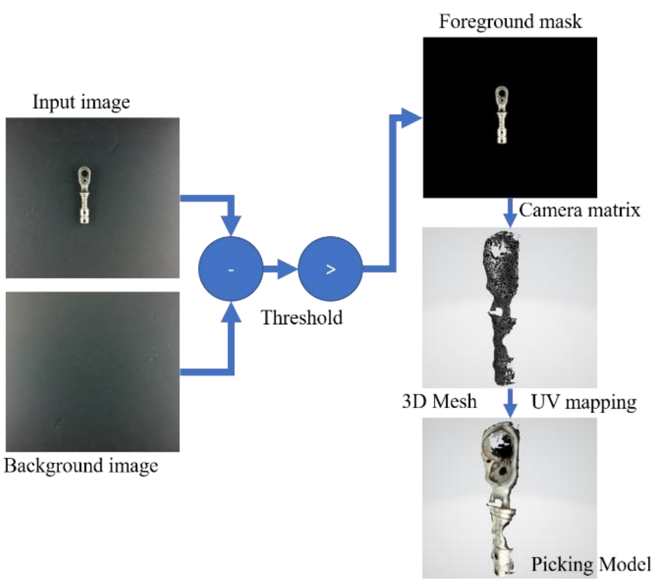

Fig. 4. Picking Modeling Pipeline.

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.

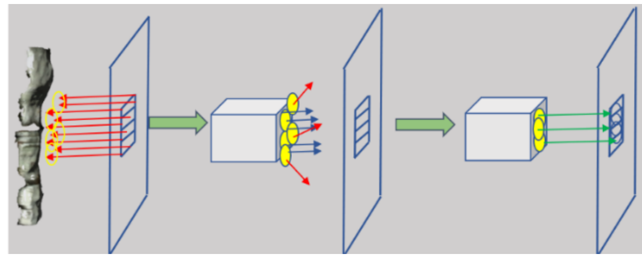

Fig. 5. Finetuning the label based on the normal of the picking point.

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. 6, can prevent grippers from successfully picking up a workpiece.

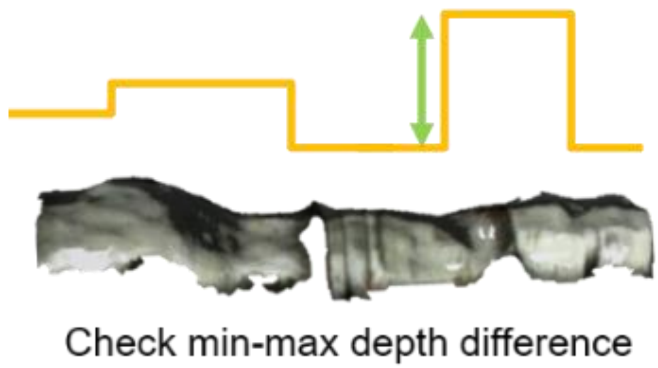

Fig. 6. Finetuning the label based on the flatness of the picking point.

Fig. 7 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.

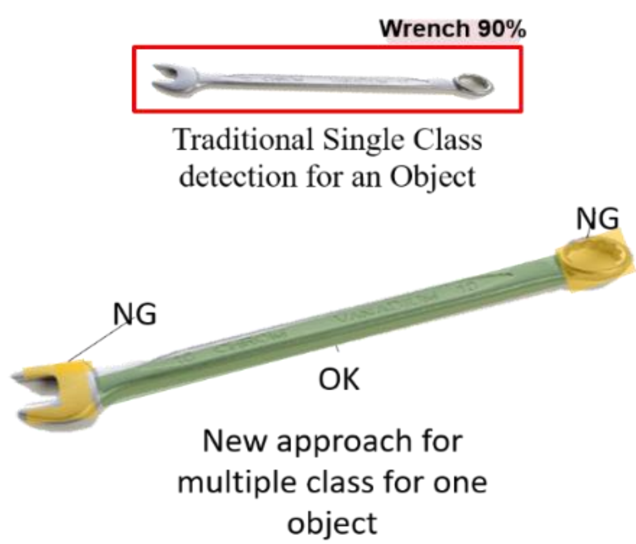

Fig. 7. Multiclass Labeling for Single Workpiece.

IV. Picking Simulations

To test picking in a simulator, we use [11, Eq.8] for simulating a stable gripping for the vacuum cup. Since we test single point picking, the equation becomes,

$$d \geq 10^{-3}\sqrt{\frac{4mt(\alpha \pm g)}{10^{-3}\pi p_v \mu}} \tag{3}$$

where d is the minimum required diameter for the vacuum cup for a successful gripping, m the mass of the workpiece, $\alpha \pm g$ the net acceleration at gripping point, p_v the pressure, μ the friction coefficient, and t the safety factor to account for non-perpendicular gripping to the ground. Note that the diameter is scaled back to meters and that pressure is taken as an absolute value.

Provided a diameter for a vacuum cup and the mass of a workpiece, we can determine the minimum pressure required for a successful grasping from

$$p_v \geq \frac{4mt(\alpha \pm g)}{10^{-3}\pi\mu(10^3 d)^2} \tag{4}$$

Likewise, provided a diameter and a maximum pressure, the maximum mass the vacuum cup will be able to hold from

$$m \leq \frac{10^{-3}\pi\mu(10^3 d)^2}{4p_v t(\alpha \pm g)} \tag{5}$$

To decrease the gap of the simulation results with the real results due to unobserved variables that affect the gripping, we use Linear Minimum Mean Squared Error (LMMSE) to reduce the MSE loss. Suppose the real result is denoted as Y and the simulated result as $\tilde{Y}$ and let them both be random variables with finite mean and variance. We want to determine a function with scalars a and b such that,

$$\hat{Y} = a\tilde{Y} + b \tag{6}$$

minimizes

$$MSE = E[(Y - \hat{Y})^2] \tag{7}$$

let us consider

$$h(a,b) = E[(Y - a\tilde{Y} - b)^2] \tag{8}$$

then, (8) is minimized if

$$a = a^* = \frac{Cov(Y,\tilde{Y})}{Var(\tilde{Y})} \tag{9}$$

and

$$b = b^* = E[Y] - aE[\tilde{Y}] \tag{10}$$

using (6), (9), and (10), allows us to re-express (4) as

$$p_v^* \geq a^* \frac{4mt(\alpha \pm g)}{10^{-3}\pi\mu(10^3 d)^2} + b^* \tag{11}$$

and using (11), (5) can be re-expressed as

$$m^* \leq \frac{a^* 10^{-3}\pi\mu(10^3 d)^2}{4(p_v - b^*)t(\alpha \pm g)} \tag{12}$$

V. Results

For the first set of experiments, we employ our AI model trained on our automatically generated dataset in the real-world RBP. We use different workpieces as specified in Tab. I and a vacuum cup with 0.04 m diameter and pressure of 80kPa. 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 average success rate is over 90% with the general workpiece.

TABLE I. Robot Picking Success Rate

Workpiece	*Mass*	*Cycle*	*Success*	*Success rate*
Candy	0.034	500	440	88%
Chewing gum	0.062	500	420	84%
Chocolate	0.050	500	480	96%
Checker Piece	0.003	500	490	98%

In the next experiment, we show the effectiveness of (3) and (10) for real-world comparison by using weights of varying masses. The minimum pressure is determined by slowly reducing the pressure of the vacuum cup until the weight is dropped. In Fig. 8, the vacuum cup holds the workpiece perpendicular to the ground, for which we set the safety factor to 1. In Fig. 9, the vacuum cup holds the workpiece parallel to the ground, thus we set the safety factor to 0.25. Further details are expressed in Table II.

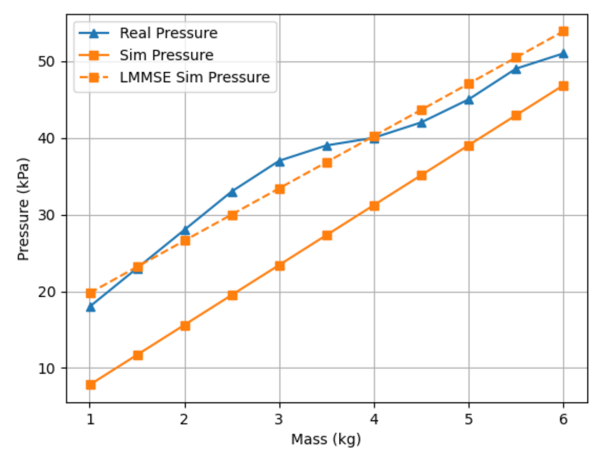

Fig. 8. Predicting the minimum pressure required for holding the workpiece with the gripper perpendicular to the ground.

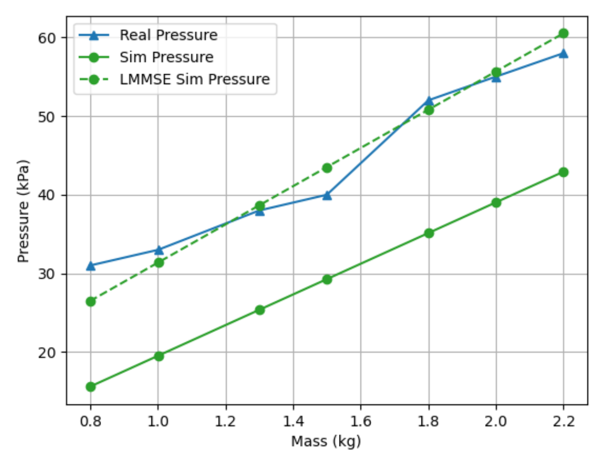

Fig. 9. Predicting the minimum pressure required for holding the workpiece with the gripper parallel to the ground.

TABLE II. SIMULATION PARAMETERS AND LOSS

μ	*t*	*d*	*a*	*b*	*MSE*
1	1	0.04	1	0	100.16
1	1	0.04	0.87	12.95	4.47
0.1	0.25	0.04	1	0	208.61
0.1	0.25	0.04	1.25	7.04	6.30

From Fig.8-9, we justify our choice of vacuum cup diameter and pressure for the earlier experiment since for workpieces of 1 kg, 25 kPa or higher is more than sufficient for grasping. Thus, for smaller masses, the statement holds.

VI. DISCUSSIONS 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.

We also provide mathematical formulas and corrections for usage inside simulations. Unfortunately, the current study is limited to vacuum cups.

As Fig. 8 and Fig.9 suggest the gap between the simulator and real-world is small however, there is still an error, albeit small. This error could be some unobserved variable as mentioned in Sec. IV or due to real-world measurement imprecision when measuring air pressure or weighing the workpiece. This measurement error may alter the result, so it is not necessarily an error from the simulation. This is a topic related to how to measure repeatability and reproducibility which we will leave as future work.

VII. 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.

We proposed 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.

We proposed mathematical formulas to analyze gripping stability by considering vacuum cup size, workpiece's mass, handling acceleration and direction, and air pressure to provide numerical analysis and to enhance realistic picking inside simulations.

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: Optimal 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.

[11] Tuleja, Peter, and L. Sidlovska. "Unilateral gripping with active vacuum suction cup Calculation of gripping force and number of suction cups." Transfer inovacii 29 (2014): 232-235.

[12] Pham, Hung, and Quang-Cuong Pham. "Critically fast pick-and-place with suction cups." *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019.

Cheng-Han (Lance) Tsai received the *M.S. degree* in *Automation and Control* from *NTUST*, *Taiwan*, in *2017*. He is currently coursing his Ph.D. degree in Power Mechanical Engineering from National Tsing Hua University, Taiwan. His current research interests and publications are in the areas of *machine learning, robotics, and digital twins*.

Jen-Yuan (James) Chang Recipient of ISPS Outstanding Contribution Award, Prof. Chang is internationally recognized for his seminal contributions to R&D and leadership in precision machines & data storage systems. His notable research contributions in rotating/transporting structure vibrations, mechatronics, wearable assistive robotic devices led to recognitions including scholarships by NRC-USAF, ASEE, and overseas institutions. He has been Technical Editor of two leading journals and has provided outstanding professional service, including Chair of ISPS Division, Vice Chair of ASME SPC and S&G Group. Currently with National Tsing Hua University, Taiwan, Prof. Chang has authored over 100 original papers, 22 patents/applications, and organized many ASME conferences.

Eduin E. Hernandez received the *B.S. degree* in Electrical and Computer Engineering Department from NCTU, *Taiwan*, in *2019.* He is currently coursing his M.S degree in Electrical and Computer Engineering Department from NYCU, Taiwan. His current research interests and publications are in the areas of *machine learning, distributed learning, approximation algorithms, and coded computing*.

Xiu-Wen You received the *B.S. degree* in Electrical Engineering Department from TKU, *Taiwan*, in *2020.* She is currently coursing her M.S degree in Electrical Engineering Department from NYCU, Taiwan. Her current research interests and publications are in the areas of *machine learning and robotics*.