IEEE Robotics & Automation Magazine (RAM) paper, presented at ICRA 2026, Vienna, Austria. Cite as RAM paper.

ROGIBOT: Development of a Dual-Arm Autonomous Container Unloading Robot System for Various Logistics Packages

Eugene Auh[1], Ilho Oh[1], Janghoon Lee[2], Nabih Pico[1], Yeongjae Park[1], Jae Hyuck Jang[1], Haneol Lee[1], Altair Coutinho[1], Jongsam Park[3], Byungtaek Jung[3], Seongyong Koo[4], Kyung-Hoon Kim[1], Hugo Rodrigue[1], and Hyungpil Moon[1]

***Abstract*— This article presents ROGIBOT, an autonomous robot designed for high-speed unloading inside standard containers. Previous unloading robots have typically been limited to handling only box-shaped packages and a small number of items at a time. These limitations hinder their applicability in general logistics environments, which involve a wide variety of package types and require high throughput. To address these challenges, we propose a novel dual-arm robot equipped with multi-functional end effectors, including a vacuum array, a two-finger gripper, and an on-hand sliding surface. The components are designed to handle the most common types of logistics packages, such as boxes, plastic pouches, and bundle sacks. In particular, the on-hand sliding surface can draw packages efficiently using the frictional force generated by a rotating elastic conveyor belt. This enables the robot to unload multiple packages consecutively at high speed by sweeping the end effectors across the pile. In addition to the hardware, we introduce a package recognition method and a finite-state-machine-based task planner that detects logistics packages and selects actions to maximize operational efficiency. In our evaluation, ROGIBOT was tested in a mock container and achieved a throughput of 2,030 pieces per hour, surpassing both state-of-the-art robots and typical human performance.**

I. Introduction

The demand for vanning and devanning tasks in logistics is increasing due to the growth of the electronic commerce market and labor shortages. In particular, devanning, also known as container unloading, is still done by human workers and is physically demanding due to the repetitive nature of lifting and moving heavy packages in a short time, and it carries a high risk of injury [1]. Thus, robotic automation of container unloading is gaining attention. Despite its necessity, the development of autonomous systems is hindered by the following major challenges:

- recognizing logistics packages of various types, sizes, and textures, and accurately estimating their poses
- handling various types of packages using a unified system
- exceeding average human performance, typically around 2,000 packages per hour by two workers loaded in a standard intermodal container[1]

Addressing these challenges requires advancements in both robotic hardware and autonomous intelligence. In this article, we present a dual-arm autonomous container unloading robot system that resolves these challenges.

Over the past decade, both academic research and industry have contributed to the development of autonomous unloading robot systems. Most existing systems have adopted a single-arm mobile manipulator combined with a telescopic conveyor belt to detect and pick carton packages using a vacuum gripper. Moreover, most existing systems have been developed for a single purpose, such as picking cartons [2], [3], [4] or grasping sacks [5], while containers typically include a variety of packages such as cartons, bundle sacks, plastic pouches, and expanded polystyrene (EPS) boxes. Furthermore, approaches using conventional vacuum or finger grippers are limited to handling a single package or a few neatly aligned boxes per cycle, which reduces overall operational efficiency. Doliotis *et al.* [2] developed a single-arm autonomous unloading robot for cartons. It segments planes from the 3D structure and extracts the vertices of the cartons to grasp two planes stably using the vacuum gripper. Jung *et al.* [3] proposed a 3 degrees-of-freedom (DOF) manipulator integrated into the conveyor belt with tilt and swivel joints. It enables the efficient pulling of aligned cartons directly on the conveyor belt; however, the automation is limited to the environment where similar carton packages are aligned in the same way. In parallel, companies

*This work was supported in part by the Technology Innovation Program (20009109, Robots that can recognize and unload unstructured and irregular packages) funded by the Ministry of Trade, Industry and Energy (MOTIE, Korea); and in part by the Technology Innovation Program (RS-2025-25449082, Development of a K-Logistics Humanoid Robot Integrated with a High-Sensitivity Robotic Hand Based on a Multimodal AI Foundation Model) funded by the Ministry of Trade, Industry and Resources (MOTIR, Korea). *(Eugene Auh and Ilho Oh contributed equally to this work.) (Corresponding author: Hyungpil Moon.)*

[1]Eugene Auh, Ilho Oh, Nabih Pico, Yeongjae Park, Jae Hyuck Jang, Haneol Lee, Altair Coutinho, Kyung-Hoon Kim, Hugo Rodrigue, and Hyungpil Moon are with the Department of Mechianical Engineering, Sungkyunkwan University, Suwon 16419, Republic of Korea {egauh, ohilho, npico}@g.skku.edu, {doug5, jjh1266, bigspirit, coutinho, khoon.kim11, rodrigue, hyungpil}@skku.edu.

[2]Janghoon Lee is with the Department of Intelligent Robotics, Sungkyunkwan University, Suwon 16419, Republic of Korea janghoon0514@g.skku.edu.

[3]Jongsam Park and Byungtaek Jung are with STC ENG Co., Ltd, Pyeongtaek 17704, Republic of Korea {sjjs56, btjung}@stceng.com.

[4]Seongyong Koo is with TES Logistics Technology Labs, CJ Logistics Corporation, Seoul 03157, Republic of Korea seongyong.koo@cj.net.

[1]*Series 1 freight containers – Classification, dimensions and rating*, ISO 668, International Organization for Standardization, Geneva, Switzerland, Jan. 2020. Available: https://www.iso.org/standard/76912.html

such as Anyware Robotics[2], Boston Dynamics[3], Kawasaki Robotics[4], and Pickle Robot Company[5] individually released single-arm unloading robots with vacuum grippers for carton packages as well; XYZ Robotics[6] released both single-arm and dual-arm loading and unloading robot systems to enhance the performance. However, existing autonomous systems are applicable only to stacked boxes, and show lower performance than human workers, which is 2,000 pieces (packages) per hour (PPH). Besides, Das *et al.* [4], in collaboration with Honeywell, proposed an autonomous unloading robot and the decision-making methods in their work. The developed robot is equipped with a wide vacuum array and a telescopic conveyor belt to pick up and pull up multiple carton boxes simultaneously. Stoyanov *et al.* [5] proposed two types of unloading robots, a novel needle chain gripper for heavy bundle sacks, and a velvet fingers gripper for various loose packages. These systems demonstrated the ability to handle packages with various types and shapes, but the cycle times exceeded one minute, which cannot satisfy the performance required by the industries. Moreover, package recognition [6], [7], [8], [9] and unloading planning [10], [11], [12] methods for automation have been proposed upon partially implemented unloading systems.

To address the shortcomings in the existing systems, we propose ROGIBOT, a dual-arm autonomous unloading mobile robot with multi-functional end effectors capable of handling a wide range of packages at high speed. The multi-functional end effectors include an on-hand sliding surface, a vacuum array, and a clamp gripper, for sweeping small packages, vacuum picking of boxes and plastic pouches, and grasping of heavy bundle sacks, respectively. In addition to the implemented hardware, we integrate the autonomous intelligence for the operation: the vision-based package recognition using deep learning and a task planner based on a finite-state machine (FSM) for high-speed unloading. The main contribution of this work is as follows:

- The proposed robotic system has demonstrated the best autonomous container unloading performance, operating over 35% faster than existing state-of-the-art (SOTA) unloading robots, handling more than 500 additional packages per hour. This improvement has been achieved through the integration of novel robotic hardware and autonomous intelligence.

[2]Anyware Robotics. "Container unloading – Anyware Robotics." Anyware Robotics. Accessed: Dec. 25, 2025. [Online]. Available: https://anyware-robotics.com/applications/container-unloading

[3]Boston Dynamics. "Stretch – Mobile Warehouse Robots | Boston Dynamics." Boston Dynamics. Accessed: Dec. 25, 2025. [Online]. Available: https://bostondynamics.com/products/stretch

[4]H. Yaname, T. Xu, Y. Takayama, K. Harukaze, and K. Morio, "Automating logistics centers by using robots," *Kawasaki Tech. Rev.*, no. 183, pp. 45–50, Mar. 2022. [Online]. Available: https://global.kawasaki.com/en/corp/rd/magazine/183/pdf/n183en09.pdf

[5]Pickle Robot. "Pickle Robot – Truck Unloading Robots – Pickle Robot." Pickle Robot. Accessed: Dec. 25, 2025. [Online]. Available: https://picklerobot.com/products

[6]XYZ Robotics. "RockyOne Mobile Manipulation Robot." XYZ Robotics. Accessed: Dec. 25, 2025. [Online]. Available: https://www.xyzrobotics.com/rocky-mmr/rockyone

The specific contributions of each component composing the system are detailed below.

- We design an end effector with an on-hand sliding surface mechanism that can unload various types of packages at high speed. The sliding surfaces use elastic rotating belts, which can generate the frictional force to draw in packages and also provide a passive compliance to compensate for errors.
- We adapt a vision pipeline that, given RGB-D images, performs instance segmentation of logistics packages and subsequently parametrizes each package in 3D space by estimating its pose and dimensions, supported by a custom dataset constructed for segmenting diverse packages under realistic container environments.
- We construct the work cycle as an FSM-based task planner based on experimentally verified heuristics. It automates the entire unloading process, including navigation of the robot and unloading actions with the multifunctional end effectors.

In this article, we describe the implementation details of the integrated system's components and evaluate its performance in a mock logistics environment. It is the best practice on autonomous unloading, which has demonstrated the best performance and has been verified in a realistic environment with various types of packages loaded inside a standard container.

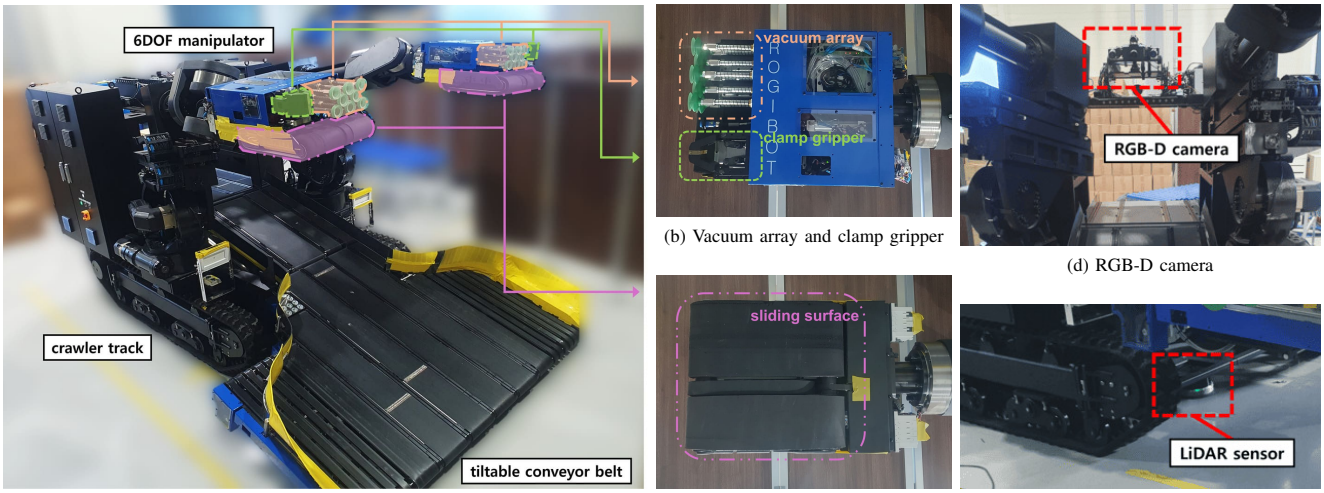
II. Unloading Robot

The unloading robot hardware is designed to meet two key requirements: handling various types of packages at high speed and moving in and out of containers. Hence, we have designed an end effector for handling both boxes and non-box packages integrated into a dual-arm mobile manipulator with a tiltable conveyor belt. In particular, the on-hand sliding surfaces installed on end effectors enable rapid unloading of various package types. In this section, we describe the robot hardware design targeting cartons, EPS boxes, plastic pouches, and bundle sacks, which comprise most of the logistics packages, and the vision sensor equipped for package recognition.

A. Dual-Arm Mobile Manipulator

The dual-arm mobile platform includes three major components: the crawler track system, the tiltable conveyor belt, and the dual 6DOF manipulators, as shown in Figure 1a. Autonomous container unloading systems must be developed as mobile robots capable of entering standard containers, which commonly have the length of 20 or 40 ft (*i.e.*, 1CC or 1AA); and internal dimensions of width 2.330m $\times$ height 2.350m. Therefore, a crawler track system is adopted for the mobile platform to enter the container and navigate on an uneven floor inside. The mobile platform is coupled with the telescopic conveyor system behind the robot to be adapted to the existing unloading environments.

A key strategy of the proposed system is to draw multiple packages simultaneously using the on-hand sliding surfaces, which causes the packages to collapse onto the forepart



(a) The dual-arm autonomous container unloading robot, ROGIBOT

(b) Vacuum array and clamp gripper

(c) On-hand sliding surface

(d) RGB-D camera

(e) 2D LiDAR sensor

Fig. 1: ROGIBOT system, its multi-functional end effector components, and sensors: (a) the robot system; (b) stretchable vacuum array for picking boxes and plastic pouches, and clamp gripper for grasping bundle sacks; (c) an on-hand sliding surface for sweeping multiple packages concurrently; (d) an RGB-D sensor between manipulators; and (e) a 2D LiDAR sensor below the tiltable conveyor belt

of the robot's conveyor system. The tiltable conveyor belt then catches and transfers the unloaded packages to the telescopic conveyor system connected to the rear part of the robot. In this process, the height of the tiltable conveyor belt is adjusted based on the assigned tasks to prevent the unexpected collapse of the lower part of the pile and minimize the impact or damage to the packages.

The dual-arm system consists of two 6DOF manipulators, each comprising five revolute joints and one prismatic joint arranged in the order of *Y-P-Pr-R-P-R*, where *R*, *P*, and *Y* denote the revolute joints in roll, pitch, and yaw directions, respectively, and *Pr* denotes the prismatic joint. In contrast to the commonly used *Y-P-P-R-P-R* configuration, which requires folding the third joint to achieve linear motion of the end effector and can result in collisions with the ceiling or floor, our design allows for safe and straightforward movement inside containers.

B. Multi-Functional End Effector

The end effectors consist of multiple components for handling various types of packages. Each of the end effectors equipped on the manipulators includes the vacuum array and the two-finger clamp gripper on one side, and the on-hand sliding surface on the other side, as shown in Figure 1b and 1c.

The on-hand sliding surface is inspired by the sliding surface gripper [13] that can manipulate contacted objects using the frictional force of the rotating belt. Drawing multiple packages using the mechanism is the core strategy of the proposed system, enabling high-speed unloading of numerous small packages. As it makes contact with the packages while the belts rotate, the resulting frictional force pulls the packages out of the container, and sweeping the manipulators can draw multiple packages onto the tiltable conveyor in a single action. This capability significantly improves unloading efficiency compared to human workers and robots that rely solely on vacuum arrays or finger grippers. Moreover, it is noted that we employ elastic belts for the on-hand sliding surfaces. The elasticity increases friction between the sliding surface and the package, while also providing passive compliance between the relatively soft packages and the position-based controlled stiff robotic manipulators. Such mechanical compliance reduces the risk of deformation or damage to the packages that may be caused by excessive contact force due to vision-based pose sensing errors [5].

Vacuum arrays and clamp grippers are for handling the following exceptional cases where the sweeping actions are unworkable. The first case occurs when packages are densely loaded in the container, preventing the end effectors from making contact with the pile. In this case, the vacuum arrays or clamp grippers are used to grasp and pull the packages located at the top-left and top-right corners out of the container, creating space for the end effectors to enter. The second case arises when the packages are laid on the container floor. In such cases, sweeping actions may collide with the tiltable conveyor belt or fail to generate enough force to lift the packages onto it. To resolve this, the vacuum arrays or clamp grippers can selectively pick up packages and place them onto the conveyor.

C. Sensors

The system uses an ENSENSO X36-5FA depth camera and a UI-5280CP-C-HQ RGB camera manufactured by IDS Imaging Development Systems to scan high-quality colored 3D scenes inside the container, as shown in Figure 1d. The structured-light 3D scanner enables accurate depth measurements in cluttered environments. Due to the sensor's limited

field of view (FOV), which can scan around 70% vertically and 110% horizontally of the container at once, we use a 1DOF actuated tilt stage to scan the entire interior of the container. Additionally, the robot equips a 2D light detection and ranging (LiDAR) sensor at the bottom part of the robot, as shown in Figure 1e, to scan and detect the inner side walls of the container and the workspace. The scanned data help the robot align with the container's side walls and approach the workspace safely by using the RANSAC algorithm.

III. Package Recognition

This section presents a vision-based method for recognizing logistics packages stacked in containers, which is essential for the operation of the autonomous system. In our system, we implement a package recognition pipeline as illustrated in Figure 2, which is based on [12]. We collected the logistics package dataset in the logistics hubs, and the vision recognition pipeline enables efficient and accurate representation of packages. The pipeline outputs oriented bounding box (OBB) parameters for box-shaped packages and segmented point cloud data (PCD) for irregularly shaped, non-box packages. During the instance segmentation stage, object masks are inferred from the input color image using deep neural networks. In the subsequent instance parametrization stage, segmented PCD are computed from the input raw points and the inferred masks. When applicable, the parameters of the OBB are estimated from the segmented PCD. This representation is beneficial because it allows the system to describe packages using compact geometric parameters, rather than relying on high-dimensional raw sensor data, thus facilitating downstream processes such as task and motion planning.

A. Logistics Package Dataset

The types of objects, stacking configurations, and the number of objects in industrial third-party logistics environments are diverse, unlike production logistics, which often contain a limited variety of goods, mostly carton boxes, arranged in a tight and neat manner. Among the few public datasets, SCD [14] contains only cartons stacked on pallets, with all instances being identical in size and shape. Moreover, the dataset only includes an average of around 20 instances per scene, whereas unloading tasks typically involve over a hundred instances in a single scene. These differences limit the generalizability of models trained on such datasets when deployed in real-world container environments.

To address this gap, a custom logistics dataset for package recognition was constructed, as described in [15], to accurately reflect real-world industrial environments. Color and depth images of packages inside containers were captured during manual unloading operations, with the scenes held stationary for the duration of the process. In the post-processing stage, visually redundant or blurred images were discarded to enhance data quality and annotation accuracy. Object instances in the remaining images were then annotated with four class labels: cartons, EPS boxes, plastic pouches, and bundle sacks, as shown in Figure 3. The final dataset consists of 1,200 images, with the distribution of each class plotted in Figure 4. As illustrated in the graph, carton and EPS boxes account for the majority of the container's contents, approximately 89%, and each scene contains an average of 83 packages.

TABLE I: Backbone comparison of SipMask in terms of accuracy (mAP) and speed (inference time) on the logistics package dataset

Backbone	mAP	mAP_{50}	mAP_{75}	Time [ms]
ResNet-50	0.614	0.793	0.678	155.5
ResNet-101	0.630	**0.805**	0.701	165.7
ConvNeXt-T	**0.681**	0.793	**0.757**	**127.0**

B. Instance Segmentation

Instance segmentation is crucial in the initial stage of the package recognition pipeline by generating object-level masks, which are used to extract point clouds for individual objects. These point clouds are then passed to the subsequent instance parametrization step. Examples of the input and output images of the process are shown in Figure 5a and 5b, respectively.

In our system, we have implemented the network based on SipMask [16]. This module classifies and segments regions of objects in the input image, such as cartons, EPS boxes, plastic pouches, and bundle sacks. Although many recent works have achieved high accuracy using transformers, we have selected SipMask as the baseline due to its lightweight architecture and fast inference time, making it well-suited for real-time applications in resource-constrained environments. For instance, the Segment Anything Model (SAM) [17], which has demonstrated high performance, uses a fixed number of queries, which inherently limits the number of instances it can detect, especially in densely packed scenes with over a hundred objects. In contrast, SipMask overcomes this limitation by leveraging FCOS [18].

Furthermore, we have replaced the backbone of the SipMask with ConvNeXt-T [19] to enhance the performance. We have assessed multiple backbones applied to the SipMask through benchmark tests on the logistics package dataset, as shown in Table I. The ResNet-50, ResNet-101, and ConvNeXt-T backbones have been evaluated, and mean average precisions (mAPs) at different intersection over union (IoU) thresholds are compared. For all three cases, the ConvNeXt-T backbone has achieved the highest mAP of 0.681 with the fastest inference time of 127.0 milliseconds. All experiments have been conducted on a system equipped with an Intel i9-12900KF CPU and an NVIDIA GeForce RTX 3090 GPU with 24 GB of VRAM.

C. Instance Parametrization

Instance parametrization is the process of obtaining shape representations of each package using instance masks and 3D point clouds. In this process, the algorithm segments the PCD and determines the lengths of the edges and the 3D pose of each box-shaped package, *i.e.*, as a 9DOF OBB. Each scene typically contains up to several hundred packages, most of

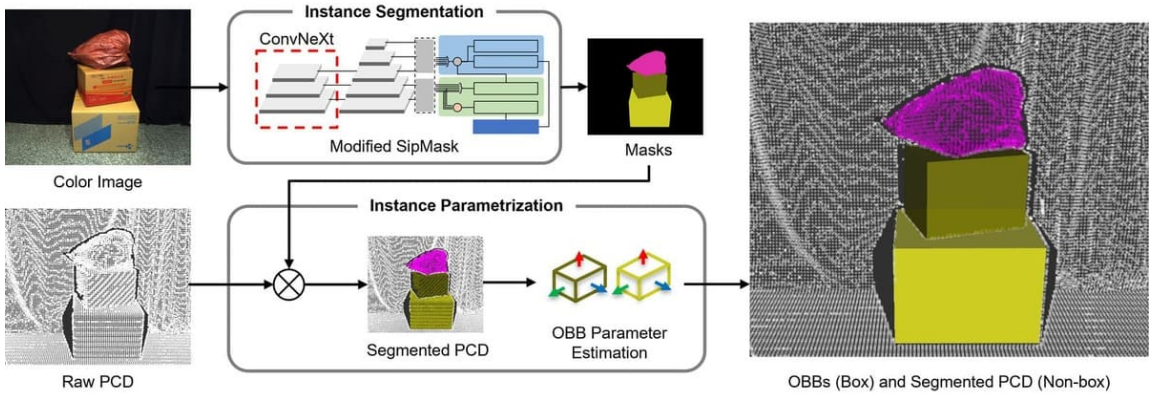


Fig. 2: Overview of the Package recognition pipeline

(a) Densely loaded cartons with various sizes and textures (b) Packages of various types (c) Collapsed packages of various types

Fig. 3: Example scenes of the logistics package dataset: (a) densely loaded cartons, (b) packages of various types, and (c) collapsed packages of various types inside containers. Each package is marked with a different color according to its type: yellow for cartons, blue for EPS boxes, green for plastic pouches, and magenta for bundle sacks.

which are box-shaped, as illustrated in Figure 5a. Therefore, representing these boxes using OBBs significantly reduces the number of parameters required to describe the scene. This approach is more efficient than representing every object using dense point clouds, especially for manipulating and estimating occluded regions. For non-box packages, such as bundle sacks and plastic pouches, we use segmented PCD to represent instances since there is no appropriate parameterization method. An example of the segmented and parameterized instances is shown in Figure 5c.

For the parameterization, we segment the point cloud using the instance masks as elaborated in Section III-B. During unloading tasks, the robot cannot observe all faces of packages, but one to three visible faces of boxes, and the occlusions complicate the parametrization process. To address this issue, the system determines the number of visible faces by iteratively extracting perpendicular planes from the segmented point cloud and calculating each edge length of the box-shaped package. If all three faces of a package are visible, the three planes in 3D space, π_1, π_2, and π_3, can be extracted iteratively using the RANSAC algorithm under the following perpendicular constraints:

$$(\mathrm{n}_1 \perp \mathrm{n}_2) \wedge (\mathrm{n}_2 \perp \mathrm{n}_3) \wedge (\mathrm{n}_3 \perp \mathrm{n}_1) \tag{1}$$

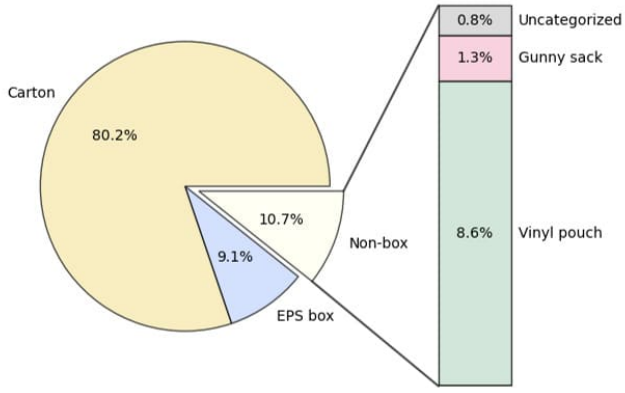
where n_1, n_2, and $\mathrm{n}_3 \in \mathbb{R}^3$ are the unit normal vectors of the extracted planes, respectively. Then, the sizes of the packages along each axis, l_1, l_2, and l_3, are computed from the inlier points, $\mathscr{P}$, identified by the RANSAC algorithm as follows.

$$l_i = \max(\mathrm{n}_i \cdot \mathscr{P}) - \min(\mathrm{n}_i \cdot \mathscr{P}) \quad i \in 1,2,3 \tag{2}$$

Otherwise, if only two faces are visible, the normal vector of the occluded face, n_3, is inferred using the perpendicular constraint between the two observed planes, π_1 and π_2, as follows.

$$(\mathrm{n}_3 \perp \mathrm{n}_1) \wedge (\mathrm{n}_3 \perp \mathrm{n}_2) \tag{3}$$

Lastly, if only a single face is visible, the direction of the package is estimated using the rotating calipers algorithm applied to the point cloud projected onto the extracted plane. In that case, only the width and height of the visible face can be calculated, and we set the length along the remaining axis (*i.e.*, the depth-wise direction) as the shorter dimension of the visible face, since its impact on the unloading task is negligible. The resultant OBBs and PCD are then utilized



Class	Quantity [PCS]	Ratio [%]
Carton	79,890	80.2
EPS box	9,051	9.1
Plastic pouch	8,550	8.6
Bundle sack	1,280	1.3
Uncategorized	839	0.8
Total	99,610	100.0

Fig. 4: Distribution of instances in the logistics package dataset

for the task and motion planning, predicting contact with the on-hand sliding surfaces and calculating the grasping pose. This compact representation allows efficient and robust downstream task and motion planning for autonomous unloading.

IV. Unloading Task Planner

In this section, we present the heuristics-based task planning method integrated into the autonomous system. It manages decision-making throughout the entire automation process, including navigation, unloading actions using multi-functional end effectors, and process termination. Based on the empirical performance of each approach, we adapt an FSM for task planning, which enables a straightforward understanding of the decision-making process.

A. Overall Flow

The unloading task planner utilizes information from the vision-based segmentation and parametrization algorithm in Section III, which provides stacked packages represented as OBBs for boxes and segmented PCD for plastic pouches and bundle sacks. We use an FSM structure for the task planner to manage the unloading process and determine the appropriate robot actions, as shown in Figure 6. After the vision sensor captures the scene (*CaptureScene*) and processes the vision recognition (*PackageRecognition*), the task planner searches the packages inside the workspace (*SearchWorkspace*).

At *SearchWorkspace* state, we check the packages within and without the workspace and find the container walls using the 2D LiDAR sensor. If packages are detected in the workspace, the task planner transitions to *Unloading* state to determine the unloading action to execute. On the other hand, if no objects are detected in *PackageRecognition* stage, we scan the surroundings with the LiDAR sensor to determine whether there are actually no objects present or if the recognition has failed. In the former case, the state can transition to *MoveForward* state to approach the distant pile, or transition to *Termination* and finish the container unloading process. In the latter case, the state returns to *CaptureScene* and *PackageRecognition* states, as shown in Figure 6, retrying to detect objects.

Subsequent to *SearchWorkspace*, the FSM states are categorized into three main groups: *Unloading*, *MoveForward*, and *Termination*, as described below.

- *Unloading* involves the actual unloading actions and includes four strategies: sweeping actions using on-hand sliding surfaces in horizontal (*HorizontalSweep*, Figure 7a) and vertical (*VerticalSweep*, Figure 7b) directions, vacuum picking (*VacuumPick*, Figure 7c), and clamp grasping (*ClampPick*, Figure 7d). The details of the unloading tasks are elaborated in the next section, Section IV-B. Once the action is determined, a motion planner calculates the trajectory of dual-arm robot (*MotionPlanning*), and the robot executes the motion (*RobotExecution*).
- *MoveForward* handles the robot's movement other than *Unloading*, such as base translation and vision sensor tilting. If the workspace near the robot is empty, it moves toward the nearest pile while aligning with the side walls, as detected by a 2D LiDAR sensor. Simultaneously, the sensor tilt is adjusted to ensure thorough inspection of the workspace. To account for the limited FOV of the vision sensor, the workspace is divided into two regions: the lower and upper sections. A threshold of 60% of the container's height is set to determine when the sensor should tilt. Initially, the vision sensor scans the upper region and tilts downward to the lower region only if the workspace above the threshold is empty. Furthermore, by verifying that the lower region is empty before the translation and inspecting the upper region afterward, we ensure that all packages in the workspace are completely unloaded.
- *Termination* indicates the completion of the process when the container is completely emptied. Once the state transitions to *Termination*, the robot gets out of the container and terminates the autonomous unloading process.

B. Unloading Tasks

For an efficient autonomous unloading process, we prioritize unloading task types to maximize the expected number of unloaded packages. As briefly described in Section II-B, the sweeping actions using the on-hand sliding surfaces enable high-speed unloading of the contacted packages. We divide the sweeping action into two types, *HorizontalSweep* and *VerticalSweep*. First, in *HorizontalSweep*, both arms approach the left and right side of the pile and then sweep the end effectors toward the center of the pile, as shown in Figure 7a. This action can consecutively unload multiple packages along the horizontal path; however, it requires empty spaces between the pile and the side walls for the

(a) Input color image (b) Instance segmentation result (c) Instance parameterization result

Fig. 5: Package recognition results: (a) an input color image, (b) resulting instance segmentation masks inferred by neural networks, and (c) resulting parameterized package instances with the planned task in three dimensions. The boxes (*e.g.*, cartons) are expressed in OBBs, and nonbox packages (*e.g.*, plastic pouches) are in segmented PCD. The 3D models of the end effectors placed in empty spaces indicate the future task to be executed, and the highlighted objects indicate the expected unloading targets by the planned task.

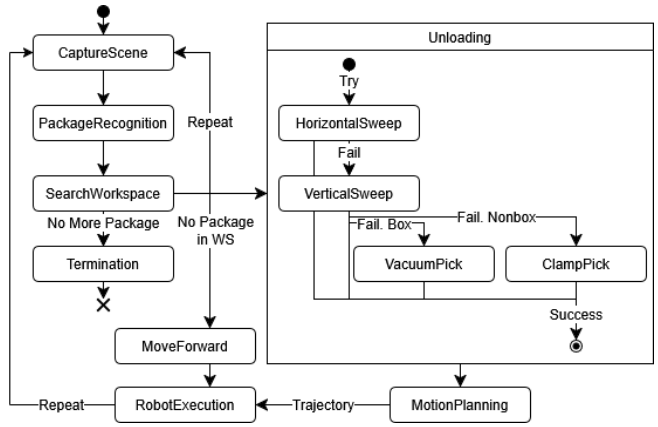


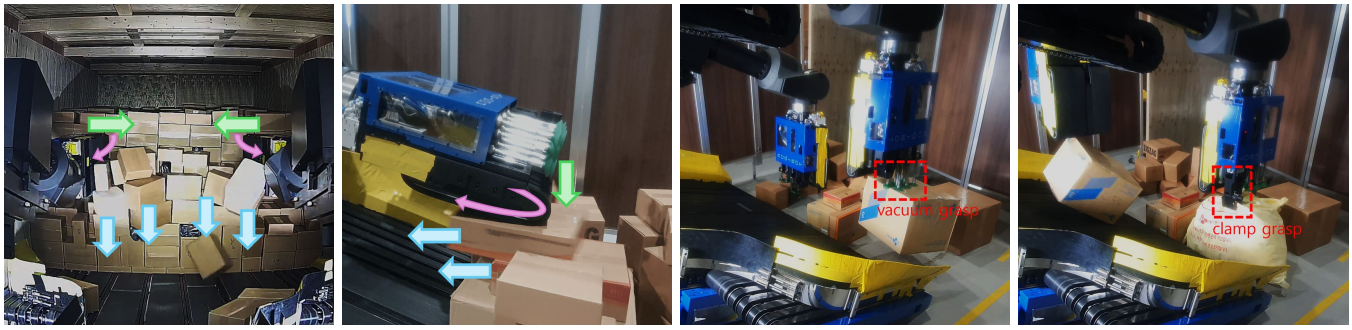
Fig. 6: FSM diagram of the developed unloading system

end effectors to enter. Second, *VerticalSweep* is the action pushing the packages downward with the rotating sliding surface as shown in Figure 7b. Additionally, *VacuumPick* and *ClampPick* actions are general picking using the vacuum arrays and the clamp grippers that can pick-and-place specific target packages, as shown in Figure 7c and 7d.

The task planner searches the empty spaces near the side walls using the occupied OBBs and PCD in the container to determine whether the end effectors can enter to execute *HorizontalSweep*, as shown in Figure 5c. If the spaces are sufficient and packages are located between, the task planner returns the decision with the searched height for motion planning. Otherwise, the task planner transitions to the state checking *VerticalSweep* motion. To perform *VerticalSweep*, the heights of the end effectors to initiate sweep motion and approaching distances are calculated from the OBBs and PCD of packages. The motion is also efficient and empties the side spaces to execute *HorizontalSweep* motion at the next work cycle. Additionally, the system adjusts the tiltable conveyor to limit the drop heights within 50 cm when executing sweeping actions, in order to prevent damage to the packages. The height can be shortened; however, it has been determined empirically, maintaining the safety and the unloading performance at the same time.

VacuumPick and *ClampPick* are used for specific cases in which the sweeping actions are not feasible. For example, picking is used when a package at the bottom might cause a collision with the conveyor, or when a heavy bundle sack makes it difficult to pull using the on-hand sliding surfaces. Once a decision is made to perform the picking action, we plan the target packages and corresponding picking postures based on the sequence planning method in [12]. For both picking actions, the OBBs and PCD are utilized to calculate the picking postures and generate safe pick-and-place trajectories. Furthermore, in cases where the packages are fully loaded up to the ceiling of the container, like in Figure 3a, the picking actions are useful to pull out packages from the corners to create space for sweeping actions. Note that *VerticalSweep*, *VacuumPick*, and *ClampPick* can be executed individually by two manipulators. For instance, the left vacuum array could pick a carton box while the right gripper grasps a bundle sack.

Based on the heuristics for maximizing performance with sweeping actions, we divide the workspace with suitable unloading task types as Figure 8, and represent the task types as states to construct an FSM-based planner, as shown in Figure 6. Within the *Unloading* state, the planner validates the feasibility and expected outcomes of each unloading task type before passing the decision to the motion planner. Note that the unloading task states in Figure 6 are for checking the feasibility of the tasks and returning the decision, not executing at the states.



(a) Horizontal sweeping (b) Vertical sweeping (c) Vacuum picking (d) Clamp grasping

Fig. 7: Unloading strategies with the multi-functional end effectors. The arrows denote the rotating direction of the sliding surfaces (purple), the direction of sweeping actions (green), and the packages' expected movements pulled out of the pile (blue).

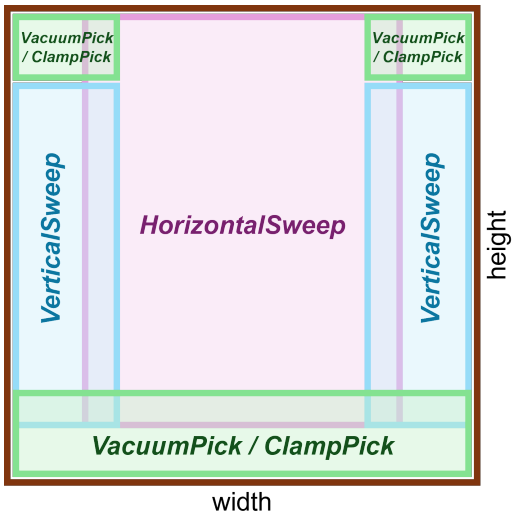


Fig. 8: Simple visualization of the effective workspace by task type in a container. The size of the container is visualized in a brown rectangle, and the workspaces of *HorizontalSweep*, *VerticalSweep*, and picking actions are marked in purple, blue, and green, respectively. Note that the picking actions are valid on the entire workspace, but are only preferred in the marked regions.

V. Evaluation

A. Vision Recognition

1) Evaluation Method: The accuracy of the vision recognition pipeline was evaluated by using the motion capture system (OptiTrack Prime^{x} 13W cameras). The evaluation metric was set to the distances between the visible plane estimated by our vision recognition method and the ground-truth plane. To calculate the distance between potentially non-parallel planes, we sampled points on the estimated plane with 10mm intervals. Then, the root mean square deviation (RMSD) between the points and the ground-truth plane has been calculated as follows:

Fig. 9: An example scene of vision recognition evaluation

$$\mathbf{RMSD}(\pi_{\text{est}}) = \sqrt{\frac{\sum_{i=1}^{N}(\mathrm{n}_{\text{GT}} \cdot (p_i^{\text{est}} - p^{\text{GT}}))^2}{N}} \tag{4}$$

where n_{GT} is the unit normal vector of the ground-truth plane, p^{GT} is an arbitrary point on the ground-truth plane, and π_{est} and p_i^{est} are the estimated plane and sampled N points, respectively. The evaluation was conducted in 50 different scenes, each containing various sizes of boxes with 10 target boxes to which motion capture markers were attached, as shown in Figure 9.

2) Result: The average of RMSDs was measured as 4.7mm with the standard deviation of 1.4mm, and the maximum value was 8.2mm. The results indicate that the plane estimated by our vision recognition pipeline can incur positional errors of up to approximately 1 cm, which didn't cause failures of the unloading actions during development and evaluation of the system.

B. Autonomous Container Unloading

1) Experimental Setup: The autonomous unloading experiments were conducted in a mock-up environment with the same dimensions as standard containers: 2.330m in width and 2.350m in height. The container was loaded

TABLE II: The autonomous unloading experiment result

Trial	Time [s]	Quantity [PCS]	Speed [PPH]
1st	1,789	1,010	2,032.42
2nd	1,792	1,025	2,059.15
3rd	1,783	989	1,996.86
Total	5,364	3,024	**2,029.53**

TABLE III: Distribution of packages for the unloading experiments. The quantities comprise all three trials.

Package type	Quantity [PCS]	Ratio [%]
Carton	2,923	96.7
EPS box	54	1.8
Plastic pouch	44	1.5
Bundle sack	3	0.1
Total	3,024	100.0

with packages of various sizes and textures, including cartons, EPS boxes, bundle sacks, and plastic pouches. The small packages, measuring W220mm×D190mm×H90mm, weighed approximately 1.5kg, while the large ones, measuring W470mm×D380mm×H340mm, weighed up to 8kg—similar to those encountered in the logistics industry. The number of packages unloaded over 30 minutes was then recorded, without any human intervention.

2) Result and Discussion: The results of the autonomous unloading experiments are presented in Table II and Figure 10. As a result, the developed system achieved an average unloading rate of 2,029.53 PPH, surpassing the performance of two typical human workers. As shown in Figure 10, it also outperforms the SOTA systems, such as Boston Dynamics' Stretch (800 boxes per hour, reported in 2021 [20]), XYZ Robotics' RockyDual (900 to 1,200 boxes per hour, reported in 2023[7]), and Pickle Robot Company's Pickle Robot (400 to 1,500 boxes per hour, reported in 2024[8]). Specifically, as listed in Table IV, *HorizontalSweep* shows the most efficient performance among the unloading task types, unloading an average of 21 packages in 26 seconds. *VerticalSweep* shows slightly lower performance compared to *HorizontalSweep*, with an average of 16 packages in 24 seconds, as its efficiency is limited by the stacked state of packages on the sides, and only one arm operates in some cases. *VacuumPick* and *ClampPick* are inefficient because they handle one or two packages at a time, but they are effective for handling heavy individual boxes and bundle sacks placed on the floor.

In our system, a few limitations can cause inefficient operations. First, if the pile collapses unexpectedly, the packages dropped on the floor have to be picked up instead of using the on-hand sliding surfaces. As shown in Table IV, picking actions are relatively inefficient because the actions can only unload one or two packages per work cycle. Therefore, although our system can handle such cases, the performance can be significantly degraded. Also, the current planning method is heuristic, relying on the current observation of the scene and empirical action performances.

VI. Conclusion

In this article, we introduced ROGIBOT, an autonomous container unloading robot system capable of handling a wide variety of logistics packages with diverse types, shapes, and textures. The key contributions of this work include the development of both hardware and software components: multi-functional end effectors integrated with a dual-arm mobile platform, a vision-based recognition algorithm for identifying various logistics packages, and an FSM-based task planning method for decision-making. In particular, the novel end effectors feature on-hand sliding surfaces that can handle a wide range of logistics packages, effectively addressing limitations of existing systems. Additionally, we collected a dataset of logistics packages and developed a neural network-based recognition pipeline. Finally, we have integrated the robot with the vision recognition and planning algorithms and conducted autonomous unloading experiments within a container environment. The results demonstrate that the proposed system outperforms both human workers and current SOTA autonomous container unloading systems. By automating the container unloading process, the proposed system can significantly reduce labor costs and enhance safety by minimizing human involvement in physically demanding tasks.

Based on the limitations discussed in Section V-B.2, potential improvements can be made in both the mechanical design and the intelligence of the system. Regarding the mechanical design, the adoption of a drawing mechanism using a tiltable conveyor belt to draw small packages on the floor, similar to [4], may improve performance by collecting packages on the floor quickly instead of picking them individually. Additionally, such a mechanism could prevent small packages on the floor from being missed by sweeping through the entire container before termination. On the intelligence side, incorporating predictions for unobserved regions and long-term task planning could further optimize the unloading process. Effective long-horizon task planning is often facilitated by accurate physical simulators—for example, to learn probabilistic models or to train reinforcement learning policies in simulation. In our container unloading setting, sweeping involves simultaneously moving and deforming surfaces with dense multi-object contact, and faithfully simulating these effects at scale remains an open research challenge, which limits simulator-centric approaches. Concretely, at the scales we target: (i) simulations of box collapse/pushing often exhibit mesh interpenetration and nonphysical rebounds, (ii) contact/friction parameters for heterogeneous materials are hard to identify and calibrate, and (iii) resolving dense, simultaneous contacts with deformable surfaces is computationally prohibitive for real-time rollouts.

[7]XYZ Robotics. "Event | Logis-Tech Tokyo 2023: XYZ Robotics' Trailer (Un)loading Solution is on its Way!" XYZ Robotics. Accessed: Dec. 25, 2025. [Online]. Available: `https://www.xyzrobotics.com/news/2023-logistech-tokyo-truck-unloading`

[8]Pickle Robot Company, *2024 Pickle Robots Unload Trucks Overview*. (Mar. 13, 2024). Accessed: Dec. 25, 2025. [Online Video]. Available: `https://www.youtube.com/watch?v=G8AH0fUIeaI`

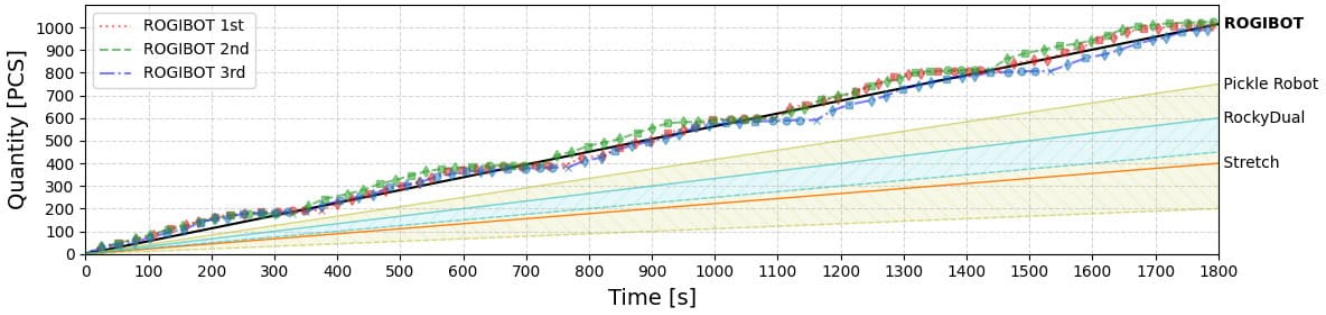


Fig. 10: Accumulated numbers of packages unloaded by ROGIBOT alongside the expected performances of the SOTA unloading robots. The task types of ROGIBOT are represented in markers: *HorizontalSweep* tasks are in squares (□), *VerticalSweep* tasks are in diamonds (◇), *VacuumPick* and *ClampPick* tasks are in circles (◯), and *MoveForward* tasks are in crosses (×). The durations comprise the partial times taken by package recognition, task planning, robot execution, and returning to the standby position. The SOTA systems' expected maximum and minimum performance levels over time are shown by the solid and dotted lines, respectively.

TABLE IV: Total numbers of executions, average quantities and times of each execution, and average efficiencies with deviations of each task type during the experiments

Task type	Total number	Quantity [PCS]	Time [s]	Efficiency [PPH]
HorizontalSweep	64	21.219 ± 9.072	26.359 ± 1.254	2886.38 ± 1219.98
VerticalSweep	101	15.723 ± 9.911	24.089 ± 3.490	2303.52 ± 1412.75
VacuumPick / ClampPick	40	1.950 ± 0.218	23.125 ± 3.628	308.11 ± 46.38
MoveForward	12	-	26.583 ± 0.640	-

ACKNOWLEDGMENTS

The dataset collection sites were provided by CJ Logistics Corporation and STC ENG Co., Ltd. The robot experiment site was provided by STC ENG Co., Ltd. We would like to thank Jaejoon Lee and Byungjin Jung from STC ENG Co., Ltd; and Seongjun Jang, Kyubum Kim, Buhm Park, and Jinkyu Chae from TES Logistics Technology Labs, CJ Logistics Corporation for their support. OpenAI ChatGPT (GPT-5 mini) was used for proofreading.

REFERENCES

[1] K. Kwak, B. Park, E. Go, C. Yoon, and K. Kim, "Rapidly spreading logistics robot applications," *J. Korea Robot. Soc.*, vol. 17, no. 4, pp. 387–396, Nov. 2022, doi: 10.7746/jkros.2022.17.4.387.

[2] P. Doliotis, C. D. McMurrough, A. Criswell, M. B. Middleton, and S. T. Rajan, "A 3D perception-based robotic manipulation system for automated truck unloading," in *Proc. 2016 IEEE Int. Conf. Automat. Sci. Eng.*, Fort Worth, TX, USA, Aug. 21–25, 2016, pp. 262–267, doi: 10.1109/COASE.2016.7743416.

[3] E.-J. Jung, S. Park, J. K. Kang, S. E. Son, G. R. Cho, and Y. Lee, "Development of a single-arm robotic system for unloading boxes in cargo truck," *J. Korea Robot. Soc.*, vol. 17, no. 4, pp. 417–424, Nov. 2022, doi: 10.7746/jkros.2022.17.4.417.

[4] M. P. Das, A. Vemula, M. Pathak, S. Aine, and M. Likhachev, "Learning optimal decision making for an industrial truck unloading robot using minimal simulator runs," *arXiv:2105.05019*, Mar. 2021, doi: 10.48550/arXiv.2105.05019.

[5] T. Stoyanov *et al.*, "No more heavy lifting: Robotic solutions to the container unloading problem," *IEEE Robot. Automat. Mag.*, vol. 23, no. 4, pp. 94–106, Dec. 2016, doi: 10.1109/MRA.2016.2535098.

[6] Z. Zhou, M. Wang, X. Chen, W. Liang, and J. Zhang, "Box detection and positioning based on Mask R-CNN [1] for container unloading," in *Proc. 2019 IEEE 4th Adv. Inf. Technol. Electron. Automat. Control Conf.*, Chengdu, China, Dec. 20–22, 2019, pp. 171–174, doi: 10.1109/IAEAC47372.2019.8997535.

[7] J. Medrano *et al.*, "Box segmentation, position and size estimation for robotic box handling applications," in *Proc. 2022 19th Int. Conf. Ubiquitous Robots*, Jeju, Republic of Korea, July 4–6, 2022, pp. 194–199, doi: 10.1109/UR55393.2022.9826264.

[8] L. Gou *et al.*, "Carton dataset synthesis method for loading-and-unloading carton detection based on deep learning," *Int. J. Adv. Manuf. Technol.*, vol. 124, no. 9, pp. 3049–3066, Feb. 2023, doi: 10.1007/s00170-022-08721-3.

[9] S. Yang, D. Li, C. Zhao, P. Wei, Y. Li, and W. Zhang, "Multi-class 4-DoF carton box detection for heterogeneous robotic container unloading," in *Proc. 2023 IEEE Int. Conf. Real-time Comput. Robot.*, Datong, China, July 17–20, 2023, pp. 1–6, doi: 10.1109/RCAR58764.2023.10249938.

[10] H. Park, G. R. Cho, E.-J. Jung, S. Park, J. Bae, and M.-G. Kim, "Reinforcement learning-based box unloading sequence planning for robotic container-unloading system," in *Proc. 2021 18th Int. Conf. Ubiquitous Robots*, Gangneung, Republic of Korea, July 12–14, 2021, pp. 371–374, doi: 10.1109/UR52253.2021.9494633.

[11] V. Giammarino, A. Giammarino, and M. Pearce, "A reinforcement learning approach for robotic unloading from visual observations," *arXiv:2309.06621*, Sep. 2023, doi: 10.48550/arXiv.2309.06621.

[12] E. Auh *et al.*, "Unloading sequence planning for autonomous robotic container-unloading system using A-star search algorithm," *Eng. Sci. Technol. Int. J.*, vol. 50, p. 101610, Feb. 2024, doi: 10.1016/j.jestch.2023.101610.

[13] J.-S. Yi, F. Yumbla, E. Auh, M. Abayebas, T. A. Luong, and H. Moon, "Passive aligning of ribbon cable in sliding surface gripper for assembly task," *J. Mechanisms Robot.*, vol. 13, no. 2, p. 025001, Apr. 2021, doi: 10.1115/1.4048915.

[14] J. Yang *et al.*, "SCD: A stacked carton dataset for detection and segmentation," *Sensors*, vol. 22, no. 10, p. 3617, May 2022, doi: 10.3390/s22103617.

[15] J. H. Kim, "Transformer based parcel instance segmentation for cluttered unloading," Master's thesis, Dept. Mech. Eng., Sungkyunkwan Univ., Seoul, Republic of Korea, Feb. 2024, doi: 10.23185/skku.000000177125.11040.0010945.

[16] J. Cao, R. M. Anwer, H. Cholakkal, F. S. Khan, Y. Pang, and L. Shao, "SipMask: Spatial information preservation for fast image and video instance segmentation," in *Proc. Eur. Conf. Comput. Vis.*, Glasgow, UK, Aug. 23–28, 2020, pp. 1–18, doi: 10.1007/978-3-030-58568-6_1.

[17] A. Kirillov *et al.*, "Segment Anything," in *Proc. 2023 IEEE/CVF Int. Conf. Comput. Vis.*, Paris, France, Oct. 1–6, 2023, pp. 4015–4026, doi: 10.1109/ICCV51070.2023.00371.

[18] Z. Tian, C. Shen, H. Chen, and T. He, "FCOS: Fully convolutional one-stage object detection," in *Proc. 2019 IEEE/CVF Int. Conf. Comput. Vis.*, Seoul, Republic of Korea, Oct. 27–Nov. 2, 2019, pp. 9627–9636, doi: 10.1109/ICCV.2019.00972.

[19] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s," in *Proc. 2022 IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, New Orleans, LA, USA, June 18–24, 2022, pp. 11 976–11 986, doi: 10.1109/CVPR52688.2022.01167.

[20] E. Ackerman, "A robot for the worst job in the warehouse: Boston Dynamics' Stretch can move 800 heavy boxes per hour," *IEEE Spectr.*, vol. 59, no. 1, pp. 50–51, Jan. 2022, doi: 10.1109/MSPEC.2022.9676361.