Digitization and Analysis Framework for Warehouse Truck Berth

Kodai Yokoyama*, Shin Katayama*, Kenta Urano*, Takuro Yonezawa*, Nobuo Kawaguchi*‡
*Graduate School of Engineering, Nagoya University
†Email: yokochin@ucl.nuee.nagoya-u.ac.jp
‡Institutes of Inovation for Future Society, Nagoya University

***Abstract*—In distribution warehouses, various forms of data transformation are underway, such as the conversion of data related to receiving and inventory management. Distribution warehouses have designated areas known as truck berths, where trucks are parked for loading/unloading. While data collection is expected to enhance work efficiency at these truck berths, there is no data regarding the status of work in progress in the unloading space, and no system is in place to analyze the work. Consequently, it is unclear which tasks are problematic and how to enhance efficiency. To address this issue, this paper proposes a framework for analyzing the loading/unloading situation at truck berths with the objective of enhancing berth utilization efficiency and work efficiency. Specifically, we employed a fixed camera for object detection through instance segmentation. This approach enables us to understand the situation of trucks, unloading spaces, and workers. The training data were annotated using a large amount of real-world data, allowing for the measurement and analysis of the efficiency of loading/unloading operations. By calculating the amount of cargo based on the floor space, we assessed the efficiency of unloading operations per truck. Through the analysis of both efficient and inefficient operations, insights for enhancing work efficiency were provided.**

***Index Terms*—Truck Berth, Deep Learning, Object Detection, Instance Segmentation**

I. Introduction

With the proliferation of Internet services, the demand for mail-order services has been on the rise. However, this growth in demand, coupled with an increase in workload and a declining birthrate along with an aging population, has led to a labor shortage problem in distribution warehouses. To address this issue, there is a need to enhance the efficiency of warehouse operations and provide adequate operational support. In warehouses, various types of information such as receiving management and inventory management are being converted into data. One crucial area in logistics warehouses is the truck berths, where trucks are parked for loading/unloading. Various efforts are underway to improve work efficiency and provide support in these berths. A significant challenge at berths is the substantial waiting time experienced by trucks. To mitigate this issue, services such as recording waiting times and implementing an advance reservation system for truck berths have been introduced. These services aim to reduce congestion and streamline truck access control, consequently enhancing operational efficiency and offering valuable operational support.

On the other hand, there is no data regarding the status of work in progress in the unloading space, and no system is in place to analyze the work.Consequently, it is unclear which tasks are problematic and how to enhance efficiency. Therefore, this paper proposes a framework for analyzing the loading/unloading situation at truck berths with the aim of improving berth utilization efficiency and work efficiency. Specifically, we employed a fixed camera for object detection [1] through instance segmentation. This approach enables us to understand the situation of trucks, unloading spaces, and workers. The training data were annotated using a large amount of real-world data, allowing for the measurement and analysis of the efficiency of loading/unloading operations. By calculating the amount of cargo based on the floor space, we assessed the efficiency of unloading operations per truck. Through the analysis of both efficient and inefficient operations, insights for enhancing work efficiency were provided.

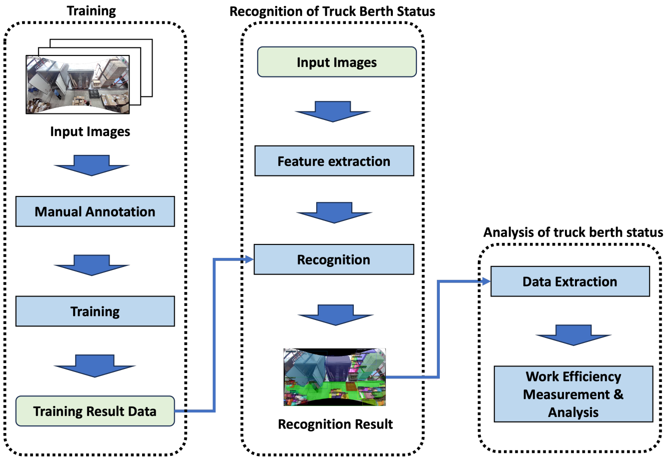

Fig. 1. Digitization flow of truck berth

The contributions of the paper are summarized as followings:

- We developed a truck berth dataset with 180k annotated objects from 5 fixed-cameras.
- We evaluated the models trained on the dataset.
- We propose a framework for digitizing and analyzing truck berth workloads.

II. Related Work

In this paper, we chose instance segmentation as we believe it is suitable for detecting individual mixed objects in a

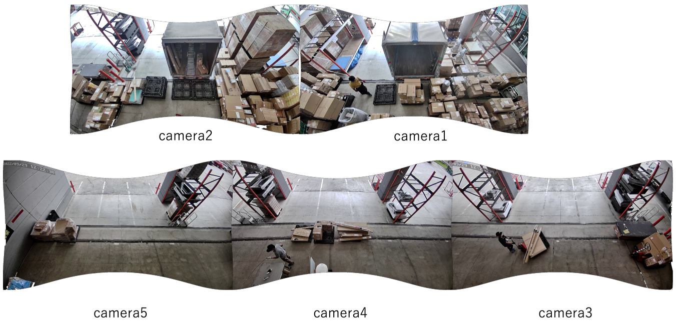

Fig. 2. Images taken in the truck berth

berth and utilizing their shape information. However, object detection has been extensively studied in the fields of pattern recognition and computer vision, and it has a wide range of applications, including object tracking. Multiple Object Tracking (MOT) is a technique that tracks multiple objects in a video, and it often employs a method called Tracking by Detection [2], which relies on object detection. This approach uses an object detection model to detect objects to be tracked in each frame of the video and then assigns the same ID to the same object from frame to frame. The object is tracked by associating IDs with the bounding boxes of the same object detected in each frame of the video. Many models supporting YOLO are used in MOT [3] [4] [5]. Therefore, in this paper, we utilized YOLO as the object detection model due to its suitability for MOT and its capability for high-speed processing.

The following studies have been conducted to improve work efficiency using object detection and segmentation in logistics warehouses. In the picking of goods, object detection and segmentation are employed to accurately identify goods and packages in warehouses, and robot arms and automated systems are controlled to pick them appropriately [6] [7].

In the transportation of goods, work efficiency has been enhanced through automation using unmanned guided vehicles and unmanned guided robots. To navigate these robots, research is also underway to derive the optimal route by recognizing shelf legs and tags through object detection [8].

III. Construction of Truck berth Dataset and its Evaluation

To digitize the truck berth, acquiring data on the movement of people and cargo is essential. Therefore, we found it effective to obtain object coordinates, shapes, quantities, and other details through object recognition using instance segmentation. However, detecting truck berth-specific objects with existing datasets proved challenging. Therefore, we endeavored to create a dataset specifically for detecting these objects. This process is illustrated in Fig. 1, where a recognition model was trained by annotating camera images of a truck berth.

A. Collecting Images

The truck berth is divided into five spaces, cameras 1-5 are installed on the ceilings of these spaces to acquire images from five different viewpoints. The images captured by cameras 1-5 are displayed in Fig. 2.

Fig. 3. Classes

TABLE I
Number of Annotations

Class Name	Train	Test	Total(Train+Test)
packed_cart	610	191	801
road_cone	374	76	450
handpallet	870	211	1081
cart	1951	499	2450
flat_trolley	323	75	398
tuck_wall	1994	508	2502
worker	2152	536	2688
pallet	28596	7061	35657
container	13643	3201	16844
packed_container	1010	297	1307
floor	1428	358	1786
cushoning	1561	378	1939
cage_trolley	1002	207	1209
cardboard	87779	22433	110212

B. Dataset

Images acquired from five cameras in the truck berth were utilized. A dataset of 179,324 annotations was created for 1,791 images. Fourteen classes were classified, as shown in Fig. 3. For annotation purposes, FastLabel was employed [9]. The number of images for training and testing was divided in an 8:2 ratio, with 1,432 images used for training and 359 for testing. The detailed number of annotations is summarized in Table I, which indicates the number of training, test, and total annotations, respectively.

C. Training

The training dataset described in the previous section was used to train the model for detection. First, we employed Detectron2 [10], a framework developed by Facebook AI Research and implemented in PyTorch. Detectron2 includes implementations of several object detection models such as Fast R-CNN, Faster R-CNN, Mask R-CNN, and others. In this paper, we utilized Mask R-CNN with ResNet-50 as the backbone architecture, which allows for instance segmentation. The next step involved training on YOLOv7 [11], a deep learning model designed for the object detection task. Both of these detection models were trained with a batch size of 8 and 400 epochs, respectively.

D. Evaluation

In this section, we provide a qualitative evaluation based on observations of actual detected images, as well as a

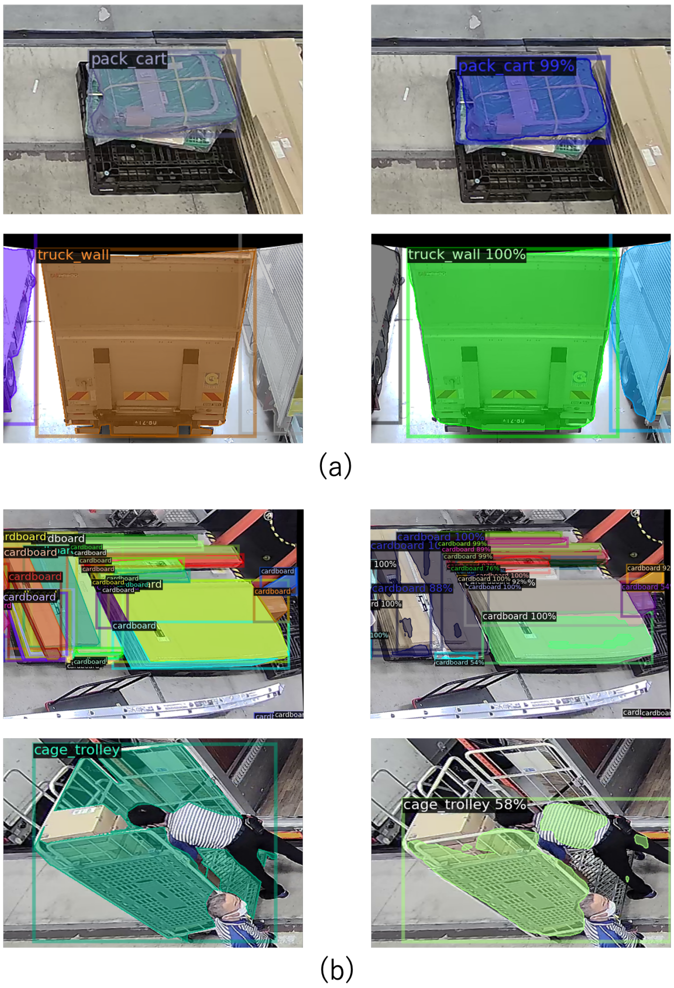

Fig. 4. Visualization of detection results (a)success (b)unsuccess

quantitative evaluation using evaluation metrics for instance segmentation and object detection.

1) Qualitative Evaluation: Below are some examples of the recognition results from the trained model. The Fig. 4 displays the visualization results of instance segmentation. In successful cases, objects stacked on top were detected with relatively high accuracy, likely because there were no overlaps or hidden areas. Moreover, trucks with closed doors were detected more accurately than those with open doors. On the other hand, instances of failure occurred when detecting objects with complex overlapping structures. This issue might be attributed to the smaller number of pixels of the object under which the cargo was placed. Another example of failure was when a worker's clothes were integrated with other objects.

2) Quantitative Evaluation: We conducted a quantitative evaluation using Average Precision (AP). Firstly, the objects with the highest accuracy were the trucks and floors. These objects have distinct characteristics; the trucks are the largest objects and often have consistent shape patterns due to their fixed position and angle. Additionally, the trucks usually do not have objects stacked on top of them, which helps maintain

TABLE II
DETECTRON2,YOLOV7.THE AVERAGE PRECISION (AP) FOR ALL CLASSES IN THE OBJECT DETECTION TASK (OBJ) AND SEGMENTATION TASK (SEG)

		AP	AP50	AP75
Detectron2	Obj	0.599	0.748	0.667
	Seg	0.432	0.652	0.478
YOLOv7	Obj	0.707	0.891	——

TABLE III
DETECTRON2,YOLOV7.THE AVERAGE PRECISION (AP) BY CLASSES IN THE OBJECT DETECTION TASK (OBJ) AND SEGMENTATION TASK (SEG)

Class Name	Detectron2 Obj	Detectron2 Seg	YOLOv7 Obj
packed_cart	**0.314**	**0.101**	**0.528**
road_cone	0.547	0.488	0.724
handpallet	0.571	0.413	0.653
cart	0.656	0.307	0.743
flat_trolley	0.434	0.133	0.595
tuck_wall	**0.958**	**0.736**	**0.989**
worker	0.576	0.505	0.465
pallet	0.513	0.223	0.682
container	0.576	0.467	0.699
packed_container	0.570	0.508	0.752
floor	0.905	0.649	0.892
cushoning	0.503	0.369	0.719
cage_trolley	0.744	0.579	0.77
cardboard	0.523	0.441	0.692

the accuracy of their mask shapes. The floor is also a large object, but its accuracy is slightly lower due to frequent shape changes caused by people and objects moving on it. On the other hand, objects with poor accuracy included packed_carts and flat_trolley. These objects have specific characteristics: they are relatively small, occur less frequently, have a smaller number of annotations, and often appear complex and overlapping. To improve the recognition of such objects piled underneath, it may be necessary to make efforts to infer chronological information.

IV. DIGITIZATION AND ANALYSIS OF TRUCK BERTH

In this chapter, we digitized and analyzed the truck berth workspace using the recognition model we developed. First, we explain the method for analyzing changes in cargo volume

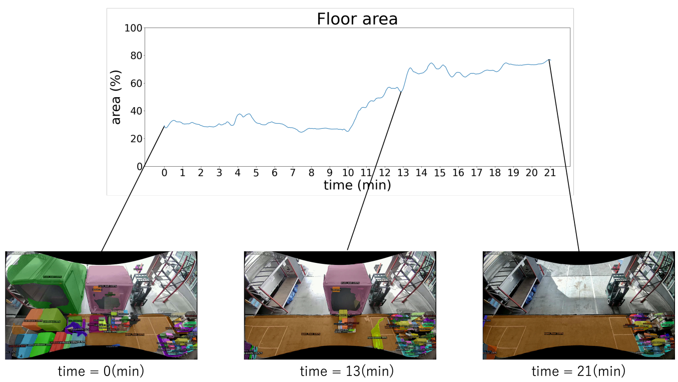

Fig. 5. Relationship between floor space change and cargo volume change

concerning floor space. Next, we measure work efficiency per truck based on cargo volume, providing insights into individual truck operations. Finally, using the measurement results and actual camera images, we analyze both efficient/inefficient operations to identify potential areas for improvement.

A. Calculation of cargo volume

Currently, detecting all cargo through instance segmentation poses challenges. As a solution, we propose a method to calculate the amount of cargo by detecting the floor area. To implement this approach, we conducted instance segmentation on each frame of a video capturing a truck berth and then calculated the number of pixels recognized as the floor. The results are displayed in Fig. 5. The graph illustrates the transition of the floor area over time, with the number of pixels representing the empty floor area normalized to the maximum value. The outputs of the instance segmentation at 0, 13, and 21 minutes are respectively displayed. An increase in floor area indicates that the loading/unloading of cargo is completed, and the floor is cleared when the truck departs. The amount of cargo can be calculated based on the changes in floor area using this method.

B. Calculation method

Perform a work efficiency analysis per truck, specifically by calculating the time required per load based on the amount of load unloaded and the time between the truck's arrival and departure. The amount of cargo is measured in terms of the number of pallets, using the fact that each pallet is equivalent to 10% of the floor space. Measurements were taken for 24 trucks, as described above.

C. Results

Efficiency is measured as the time taken per pallet. The average work efficiency time was 398.6 seconds, with the most efficient work taking 157.7 seconds and the least efficient work taking 918.0 seconds. The analysis of the results with good and bad measurements was conducted using actual camera images. The most efficient and least efficient results showed a time difference of about 5.2 times. One of the contributing factors to this difference was whether all the cargo was initially placed on the pallet. When the cargo was already on the pallets, it could be smoothly moved using hand pallets, resulting in efficient loading/unloading. On the other hand, when the cargo was loaded separately, transferring each item to pallets took considerable time. These results highlight the importance of efficiently placing as much cargo as possible on pallets. Another factor impacting efficiency is the placement of unloaded loads in different locations. Some are carrying pallets to the back, while others others just unload the pallets and leave. The difference in pallet placement can significantly affect work time and lead to confusion among warehouse workers. Therefore, unifying the placement of pallets is also considered necessary to improve work efficiency.

V. Conclusion

In this paper, we proposed a framework for digitizing and analyzing the truck berth workspace, aiming to enhance berth utilization and work efficiency through the digitization of truck berth workspaces. To achieve this goal, we created a truck berth dataset dataset with 180k annotated objects and trained an instance segmentation model. Utilizing this model, we conducted an analysis of work efficiency per truck by detecting floor space as a part of the digitization and analysis of the berth workspace.

For future perspectives, it's important to note that the work analysis was limited to cases where a single truck was parked. However, we are considering expanding the analysis to scenarios with multiple parked trucks. In such cases, changes in floor space alone may not be sufficient to identify which cargo belongs to each truck. To address this, cargo tracking is being considered. Consequently, improving the detection accuracy of cargo objects, such as cardboard boxes, is essential. We are also exploring options like revising the labeling approach and exploring other object detection models to enhance the analysis.

Acknowledgment

This work was supported by JST, CREST Grant Number JPMJCR21F2, NICT(22609), JSPS Grant Number JP22H03696, JP22K18422, NEDO JPNP23003 and TRUSCO NAKAYAMA Co., Ltd.

References

[1] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 39, No. 6, pp. 1137–1149, 2017.

[2] Zdenek Kalal, Krystian Mikolajczyk, and Jiri Matas. Tracking-learning-detection. *IEEE transactions on pattern analysis and machine intelligence*, Vol. 34, No. 7, pp. 1409–1422, 2011.

[3] Nir Aharon, Roy Orfaig, and Ben-Zion Bobrovsky. Bot-sort: Robust associations multi-pedestrian tracking. *arXiv preprint arXiv:2206.14651*, 2022.

[4] Mikel Broström. Real-time multi-camera multi-object tracker using yolov7 and strongsort with osnet. https://github.com/mikel-brostrom/Yolov7_StrongSORT_OSNet, 2022.

[5] yolov7-deepsort-tracking. https://github.com/deshwalmahesh/yolov7-deepsort-tracking.

[6] Rico Jonschkowski, Clemens Eppner, Sebastian Höfer, Roberto Martín-Martín, and Oliver Brock. Probabilistic multi-class segmentation for the amazon picking challenge. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1–7, 2016.

[7] Andy Zeng, Kuan-Ting Yu, Shuran Song, Daniel Suo, Ed Walker, Alberto Rodriguez, and Jianxiong Xiao. Multi-view self-supervised deep learning for 6d pose estimation in the amazon picking challenge. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1386–1383, 2017.

[8] Yiyou Sun, Tonghua Su, and Zhiying Tu. Faster r-cnn based autonomous navigation for vehicles in warehouse. In *2017 IEEE International Conference on Advanced Intelligent Mechatronics (AIM)*, pp. 1639–1644, 2017.

[9] Fastlabel. https://fastlabel.ai/.

[10] Yuxin Wu, Alexander Kirillov, Francisco Massa, Wan-Yen Lo, and Ross Girshick. Detectron2. https://github.com/facebookresearch/detectron2, 2019.

[11] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv preprint arXiv:2207.02696*, 2022.