Digitization Methods for a Logistics Warehouse Towards Digital Twin-Driven Optimization

Nobuo Kawaguchi
Inst. of Innov. for Future Society
Nagoya University
Nagoya, Japan
kawaguti@nagoya-u.jp

Yusuke Asai
Grad. Sch. of Eng.
Nagoya University
Nagoya, Japan
asayu@ucl.nuee.nagoya-u.ac.jp

Kazuma Kano
Grad. Sch. of Eng.
Nagoya University
Nagoya, Japan
kazuma@ucl.nuee.nagoya-u.ac.jp

Kairi Takaki
Grad. Sch. of Eng.
Nagoya University
Nagoya, Japan
kairi@ucl.nuee.nagoya-u.ac.jp

Yuki Mori
Grad. Sch. of Eng.
Nagoya University
Nagoya, Japan
ymori@ucl.nuee.nagoya-u.ac.jp

Yuma Suzuki
Grad. Sch. of Eng.
Nagoya University
Nagoya, Japan
chanyou@ucl.nuee.nagoya-u.ac.jp

Kisho Watanabe
Grad. Sch. of Eng.
Nagoya University
Nagoya, Japan
kisho@ucl.nuee.nagoya-u.ac.jp

Yuki Gushi
Grad. Sch. of Eng.
Nagoya University
Nagoya, Japan
gushi@ucl.nuee.nagoya-u.ac.jp

Shin Katayama
Grad. Sch. of Eng.
Nagoya University
Nagoya, Japan
shinkatayama@nagoya-u.jp

Kenta Urano
Grad. Sch. of Eng.
Nagoya University
Nagoya, Japan
urano@nagoya-u.jp

Takuro Yonezawa
Grad. Sch. of Eng.
Nagoya University
Nagoya, Japan
takuro@nagoya-u.jp

Shintaro Hashiguchi
Logistic Platform Development and Planet AICHI's Preparation Office
TRUSCO NAKAYAMA Corp.
Tokyo, Japan
shintarou.hashiguchi@trusco.co.jp

***Abstract*—The rapid growth of e-commerce, rising consumer expectations, and labor shortages in Japan pose challenges for logistics warehouse optimization. While automation has advanced outbound operations, inbound processes remain inefficient. This study presents a comprehensive approach to digitizing and optimizing large-scale logistics warehouses, with a case study at TRUSCO NAKAYAMA Corp. in partnership with Nagoya University. Key technologies include a large-scale camera array, multi-camera object tracking, and smartphone-based task estimation. Our contributions include real-time tracking of personnel and packages, cooperative annotation for improved object recognition, and synthetic data augmentation. Additionally, truck berth analysis and indoor localization enhance operational efficiency. To optimize worker shifts and warehouse layout, we apply Factorization Machine Quantum Annealing (FMQA), achieving a 37.4% reduction in lead times and a 14.3% decrease in labor hours. A visualization tool enables warehouse operators to make data-driven decisions. This research demonstrates the potential of digital transformation in logistics and provides a scalable framework for broader industry adoption.**

Keywords*— *logistics, warehouse, image recognition, object tracking, quantum annealing, simulation

I. Introduction

The rapid growth of e-commerce, rising consumer expectations for faster deliveries, and labor shortages in Japan pose significant challenges for optimizing logistics warehouse efficiency. Despite technological advancements, automation efforts have mainly targeted outbound processes, leaving inbound operations largely inefficient.

TRUSCO NAKAYAMA Corp. is a specialized trading company that wholesales machinery, tools, logistics equipment, and environmental safety products for use in factories and construction sites. The company handles approximately 4.5 million items from over 3,500 manufacturers worldwide and manages an inventory of approximately 600k items across 28 logistics centers throughout Japan. To establish an advanced logistics hub, TRUSCO NAKAYAMA entered into a comprehensive partnership agreement with Nagoya University in 2021 and has been conducting joint research.


Fig.1. Perspective photo of target logistics warehouse

Fig.1 shows perspective photo of our target logistics warehouse(1st floor of Planet Tokai) which is one of the logistics centers of TRUSCO NAKAYAMA. This paper introduces various efforts related to the construction of a digital twin for large-scale warehouse optimization. While existing research has typically addressed specific components of warehouse digitization, such as object tracking, image recognition, or workforce scheduling independently, there has been limited effort in developing a comprehensive, fully integrated system that spans from real-time data acquisition to operational optimization. Specifically, most prior studies do not evaluate their proposed methods in actual large-scale warehouse environments over extended operational periods. To bridge this gap, our study proposes and implements an end-to-end

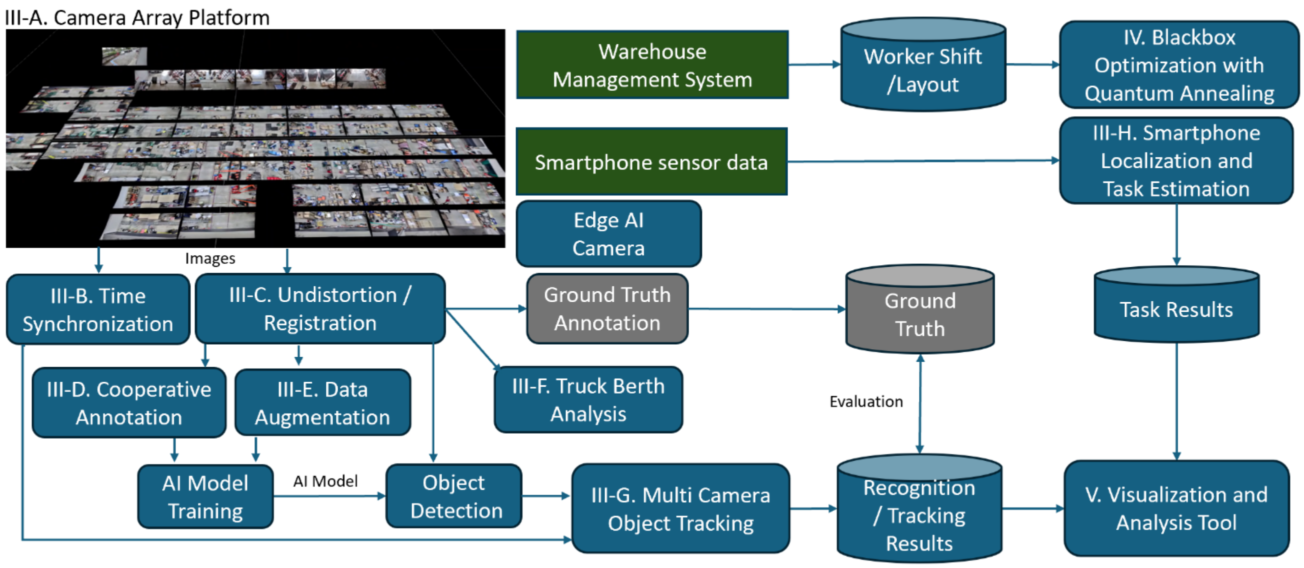

Fig.2. Data/process flow of digitization methods for logistics warehouse

warehouse digitization and optimization system, covering all aspects from data collection (camera array and smartphones) through integrated analysis (multi-camera tracking and task estimation) to operational optimization (quantum annealing-based shift scheduling). A key novelty of our work is that we have successfully deployed and continuously operated this comprehensive digital twin-driven system in a real-world large-scale logistics facility, TRUSCO NAKAYAMA Corp.'s logistics center, providing practical validation and insights beyond laboratory conditions.

II. Related Work

Several studies have explored logistics warehouse optimization to enhance operational efficiency. For instance, [1] introduces a wide range of research and development efforts from the perspective of the Sustainable Development Goals (SDGs). Specific approaches include reducing total costs and improving efficiency through IoT utilization, visualizing inventory using Augmented Reality (AR) technology, optimizing processes with QR codes and RFID, leveraging AI and chatbots for data collection and analysis, and employing automated transport systems.

Furthermore, [2] has released datasets for image recognition aimed at improving Pick & Place efficiency, such as those used in the Amazon Picking Challenge [3]. Additionally, [4] focuses on achieving autonomous forklifts in warehouses, collecting vast amounts of warehouse images, including those of people and pallets, to build a pallet database and enhance efficiency through deep learning. In [5], the authors highlight the importance of positioning systems in warehouse optimization and propose a testing and evaluation framework for their implementation. [6] has released a large-scale multimodal dataset, OpenPack, for packaging work recognition in warehouses.

These studies demonstrate that numerous challenges remain in warehouse efficiency optimization, and various research efforts continue to address specific issues. We strongly believe that data-driven analysis is essential for warehouse optimization.

In this context, [7] proposes an algorithm model that combines YOLOv5[8] and DeepSORT[9] to enable object tracking, with the goal of digitizing logistics warehouses. The evaluation results indicate great potential for real-world warehouse applications, yet practical implementation in actual warehouse operations has not yet been realized.

This study aims to digitize warehouse operations by recognizing and tracking large-scale packages and people, as well as identifying their behaviors in real-world warehouse environments.

III. Digitization Methods for Logistics Warehouses

Fig.2 illustrates data and processes flow of the proposed digitization methods for logistics warehouses, which are addressed in this study. The following sections provide a sequential explanation of each component.

A. Camera Array Platform

In this study, the primary target area is the first-floor of the warehouse, which has a ceiling height of approximately 6 meters. Incoming packages can sometimes be stacked over 2 meters high, making it difficult to capture all activities using overhead cameras positioned at an oblique angle, as occlusion occurs, preventing a complete record of warehouse operations.

To overcome this limitation, we installed more than 60 wide-angle network cameras (H.View HV-800G2A5, priced at approximately $120 each) on the ceiling of the first-floor warehouse, as shown in Fig.1. These cameras are oriented vertically downward to record the movements of workers and pallets/parcels. Each camera transmits H.264 video streams via RTSP, which are then stored in an on-site storage system.

While Edge AI[10,11] implementation is planned for future deployment, we initially opted for full video storage to ensure comprehensive data collection. As a result, the recording of more than 80 camera feeds generates approximately 1.3 terabytes of video data per day, totaling over 500TB in 20 months.

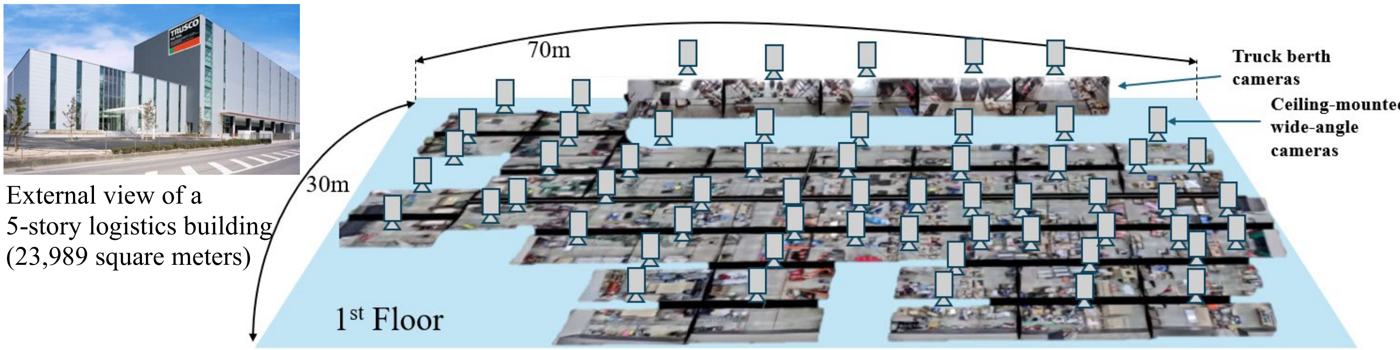

Fig.3. External view and camera array platform of the target logistics warehouse (part of 1st floor).

B. Video Time Synchronization

Given the large-scale deployment of low-cost surveillance cameras, precise hardware-based synchronization among cameras is not available. Instead, NTP (Network Time Protocol) is employed for time correction. The video feed is recorded at 5 frames per second (fps), and each image contains a timestamp with a resolution of one second.

To ensure proper time synchronization across frames, we utilize optical character recognition(OCR) techniques, where the first detected timestamp change serves as a reference point for frame alignment. The NTP synchronization is scheduled to run once per hour; however, failures in synchronization occasionally occur, resulting in timestamp discrepancies of several seconds. In such cases, manual adjustments are performed as needed.

Future research should focus on developing robust synchronization techniques to mitigate incidental time drift and enhance the reliability of time alignment across multiple camera feeds.

C. Undistortion, Registration and Stiching of Camera Images

Due to installation constraints, we have placed wide-angle cameras with a 110-degree field of view (FOV) unevenly on the ceiling of the warehouse. Consequently, distortion correction (undistortion) and registration of camera positions are required.

For wide-angle camera undistortion, we initially utilized OpenCV's fisheye camera model. However, the significant distortion posed a challenge. Therefore, we conducted a study on undistortion methods and adopted the Double Sphere Model [12], which offers a well-balanced trade-off between speed and performance. Nevertheless, the issue of camera distortion became apparent only after fixing the cameras to the ceiling, leaving unresolved challenges in camera calibration. As an important insight gained through experience, we emphasize that selecting the appropriate camera model and verifying its parameters should be carefully conducted before securing the cameras in place.

The distortion-corrected camera images are used for object recognition, and we also implemented image stitching to monitor the overall floor conditions. Since each camera covers a different area, registration and masking of positions are

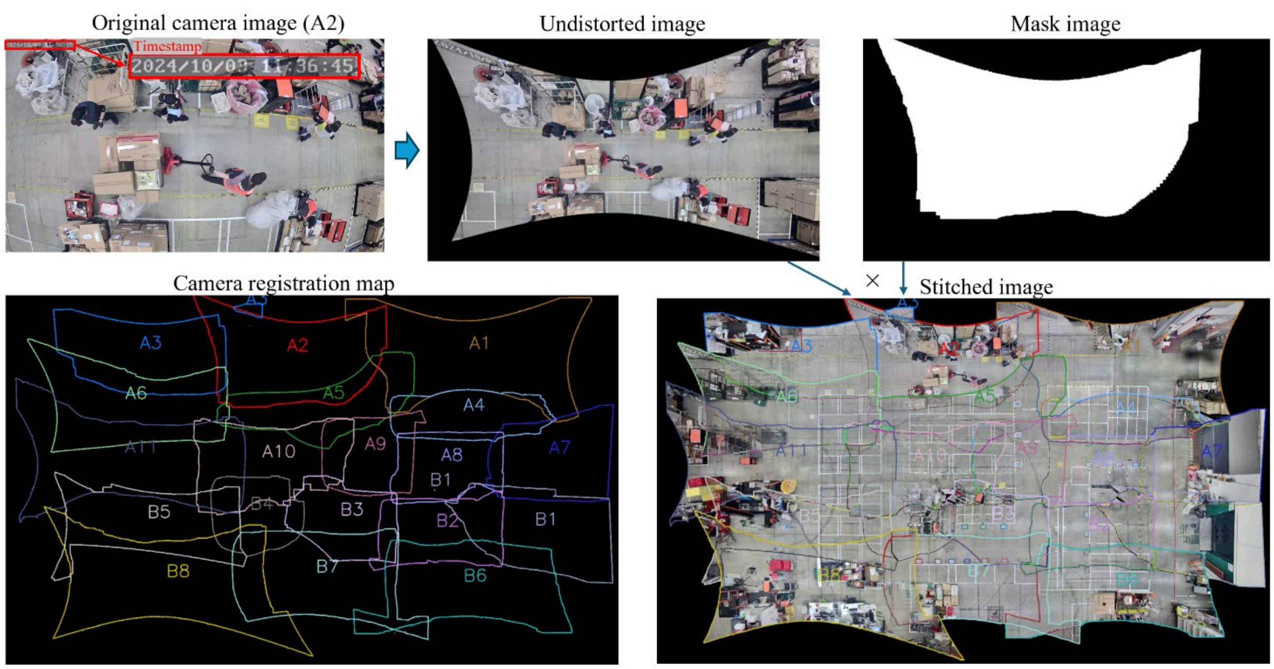

Fig.4. Camera image undistortion and stitching

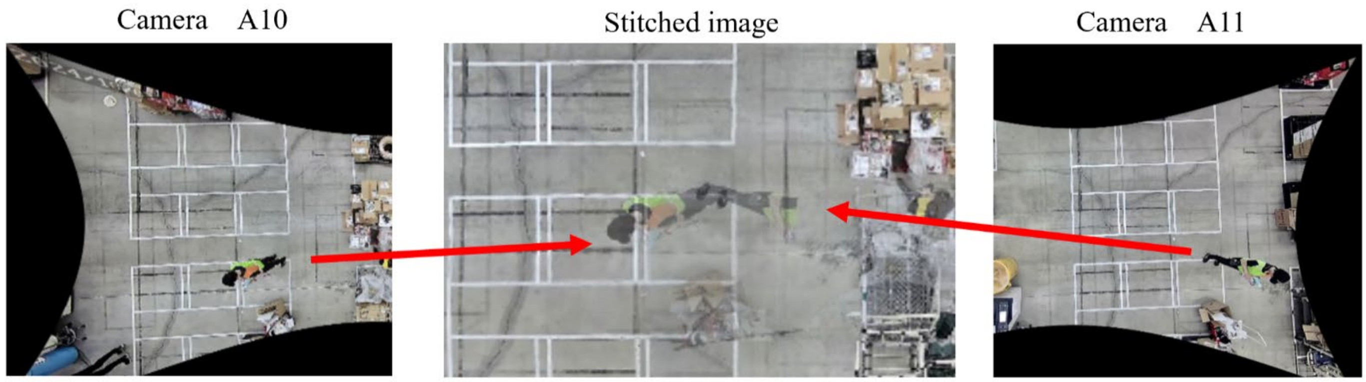

Fig.5. Stitching with alpha blending

necessary. For camera position registration, we utilized a colorized point cloud of the warehouse floor acquired using Leica's BLK2GO [13] and performed SuperPoint[14] and LightGlue feature matching [15] between the images. Fig. 4 presents the captured camera image (only A2), the undistorted camera image, the mask, the map after registration, and the results of stitching. During the stitching process, alpha blending was applied to overlapping areas of the masked images. As a result, workers captured by multiple cameras are blended and displayed from different angles. If stitching were performed without overlapping camera images, areas further from the camera center might not properly represent objects above the floor. Thus, alpha blending plays a crucial role in ensuring seamless image integration. Fig.5 shows the effect of alpha blending which results duplicated worker image on the stitched image.

D. Cooperative Annotation

To recognize workers, tools, and packages within a logistics warehouse using multiple cameras, object detection based on deep learning proves to be highly effective. In this study, we employ Detectron2 [16] as the object detection model. Given that the cameras are installed on the ceiling, leading to a unique field of view, and that the logistics warehouse itself represents a specialized environment, pre-trained models of Detectron2 do not yield satisfactory performance. Consequently, custom annotations and trainings are required.

Furthermore, the distortion introduced by wide-angle cameras results in various patterns for workers, tools and packages, making exhaustive manual annotation excessively time-consuming process. To address this challenge, we take advantage of the fixed camera setup and construct a framework [17,18] that assists annotation by extracting only moving objects using optical flow and leveraging their similarity.

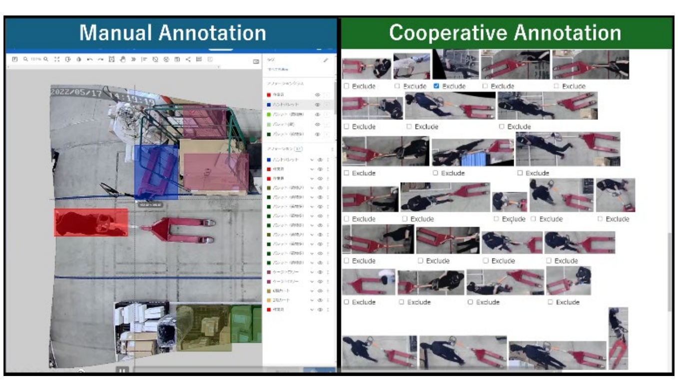

Fig.6. Manual and Cooperative Annotation

Details of the framework can be found in the full paper. Without the framework, individual objects had to be manually annotated and labeled as shown in Fig. 6 (left). With the framework, similar images are displayed using a tool, allowing annotators to exclude only incorrect images, as shown in Fig. 6 (right). This significantly accelerates the annotation process.

Through a technical survey of individual components, we determined that RAFT [19] is effective for optical flow-based moving object segmentation. Moreover, SimSiam [20] is utilized for evaluating the similarity of segmented objects, UMAP [21] for dimensionality reduction, and K-Means for clustering. As a result, the proposed framework achieves a 98% reduction in annotation time while maintaining the same annotation volume. However, collaborative annotation may result in the annotation of extraneous regions, which poses a challenge to recognition performance.

E. Data Augmentation using Synthetic Images

In cooperative annotation, unintended annotation errors have been observed to degrade recognition performance. One primary reason for this degradation is the inclusion of training images that lack annotations for detectable objects, despite their presence in the scene.

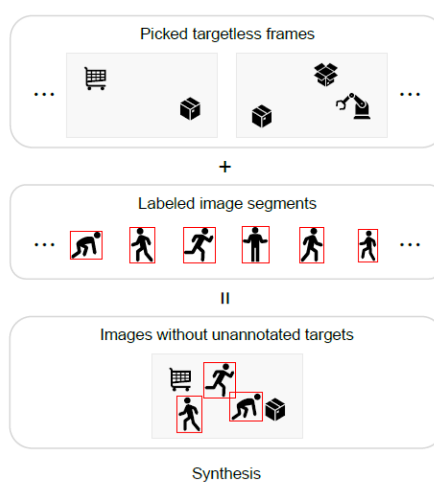

Fig.7. Synthetic image generation

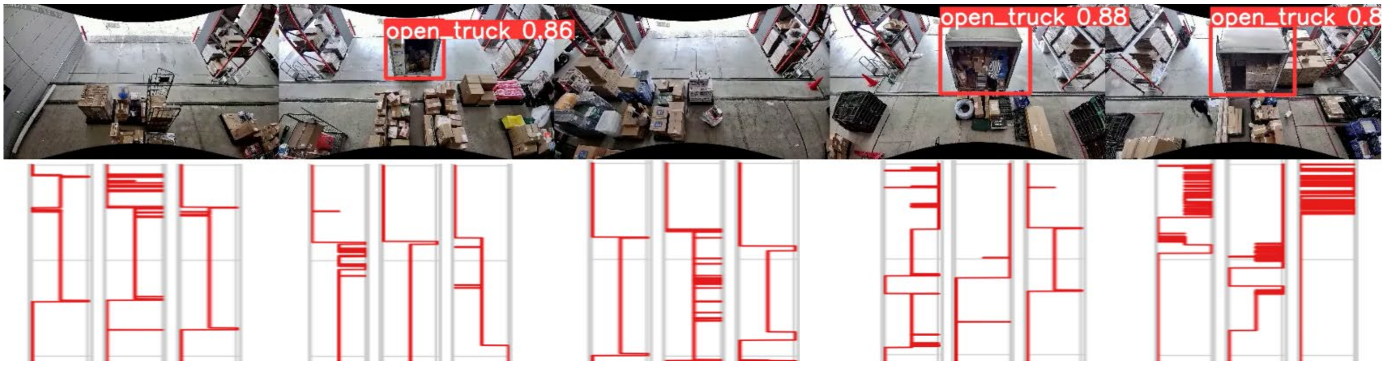

Fig.8. Recognition and visualization of truck berth status. Bottom graph is a time series of status (empty, close_truck and open_truck).

To address this issue and construct accurate and diverse annotation datasets, we propose a data augmentation method [22] based on synthetic image generation. The overview of this method is illustrated in Fig. 7. Specifically, our approach synthesizes natural-looking images without unannotated objects by overlaying labeled foreground segments onto background images that do not contain any detection targets. These background images are essential to ensure that no unannotated objects remain in the synthesized images.

To efficiently extract background images without target objects, our method leverages the motion information provided by RAFT, assuming that target objects exhibit movement. Concretely, background frames are collected from randomly sampled frames where RAFT did not detect any motion. The inclusion of unrelated objects within these frames contributes to dataset diversity. Although human verification of a subset of frames for each camera is required, the workload is significantly reduced compared to manual annotation. The more background frames are prepared, the greater the diversity of the generated images; however, this also increases the associated workload.

Furthermore, by utilizing the ceiling-mounted camera setup, the method preserves the original positions of the segmented foreground objects, ensuring the generation of realistic images. Evaluation using video footage from a logistics warehouse confirmed that improving dataset reliability directly enhances model performance.

F. *Truck Berth Analysis*

In logistics warehouses, truck berths are designated areas where trucks enter and exit for loading and unloading cargo. The target logistics warehouse in this study has 15 truck berths, with five fixed-point cameras installed, each covering three berths.

To analyze berth occupancy, we first developed a simple recognition model to detect the presence of trucks and the status of their rear doors (open or closed). As shown in Fig. 8, we constructed a system to track the status of each truck over time. While warehouse operators had an intuitive understanding of truck berth usage, they lacked a systematic method for visualizing and quantifying berth utilization. To obtain more detailed insights into cargo handling activities, an advanced cargo recognition model is required. Since the truck berth cameras are installed at an angled perspective, conventional object detection models can be applied. However, the diversity of cargo types poses a challenge for accurate recognition. To address this, we employed instance segmentation to identify individual cargo items and performed detailed annotations for model training [23].

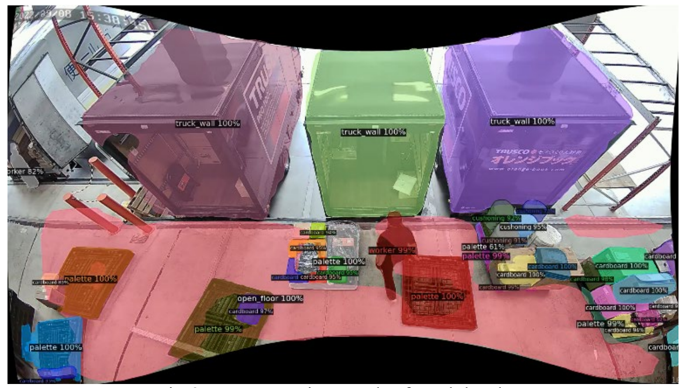

Fig.9. Segmentation result of truck berth

Specifically, annotations were conducted on 2,035 images covering 14 target classes, resulting in a total of 193,357 labeled instances. The segmentation was performed using Detectron2, with the results illustrated in Fig. 9. As detailed in the paper [23], we further analyzed the work environment by utilizing the floor area extracted from the segmentation results to assess operational efficiency.

G. *Multi-Camera Tracking*

To effectively utilize the AI model trained with a large number of annotations obtained from subsections D and E, it is necessary to integrate recognition results obtained from video

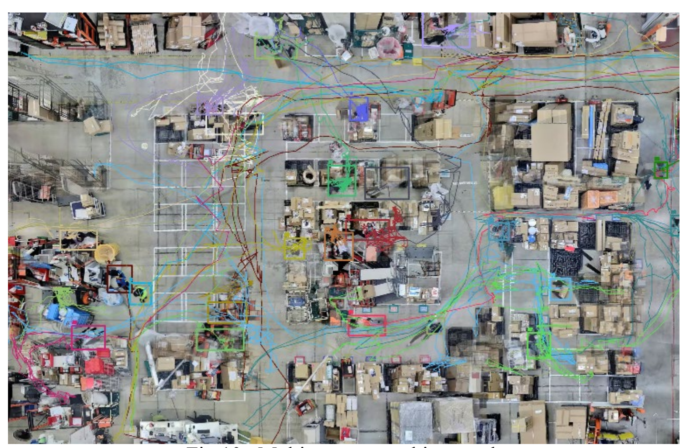

Fig.10. Multi-camera tracking result

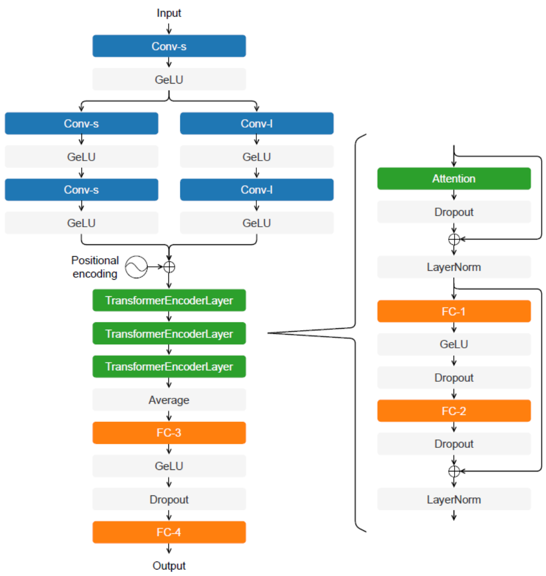

Fig.11. DualCNN-Transformer

image across multiple cameras. For intra-camera tracking, existing technologies such as TrackFormer [24] can be employed. However, inter-camera tracking presents unique challenges due to the use of wide-angle cameras, making direct application of conventional methods ineffective.

To address this issue, we have enhanced ByteTrack [25] by refining the selection method for bounding boxes in recognition results, thereby developing a practical tracking approach. The integrated recognition results of the current warehouse environment are used to visualize worker movement trajectories, as illustrated in Fig. 10.

H. Indoor Localization and Task Estimation

In parallel with warehouse digitization using cameras, we also utilize IMU sensors embedded in workers' smartphones to conduct indoor positioning and task estimation. Unlike simple pedestrian movement, warehouse operations involve various tasks, making conventional pedestrian dead reckoning ineffective. To address this challenge, we integrate an end-to-end pedestrian velocity estimation method [26] with a gait-robust orientation estimation technique [27]. Additionally, by incorporating photovoltaic-powered BLE beacons, we employ a DualCNN-Transformer model, as illustrated in Fig. 11, to achieve high-accuracy indoor positioning within the warehouse [28].

Warehouse workers primarily engage in three types of tasks: inspection, sorting, and transportation. To quantify the distribution of these activities, we conducted task estimation using smartphone-based motion analysis[29]. Specifically, as shown in Fig. 12, six-axis acceleration and angular velocity data were collected from smartphones attached to workers' waists. Using a 5.12-second sliding window and logistic regression based on the features listed in Table I, we achieved an F1-score of 0.83 for task estimation. Sample results of task estimation are shown in Fig. 13.

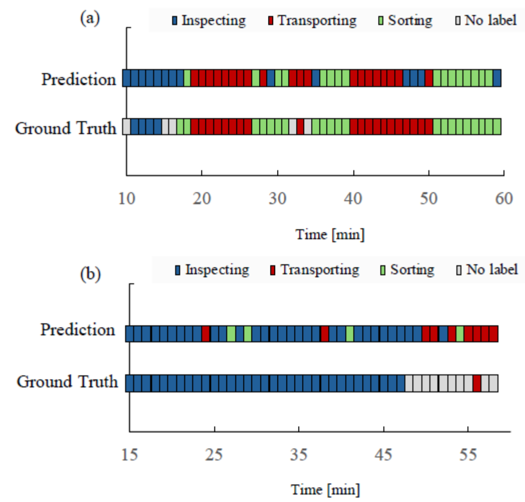

Fig.13. Sample results of task estimation.

TABLE I. FEATURES USED IN LOGISTIC REGRESSION

	Features
Time domain	Mean, Standard deviation, Maximum, Minimum, IQR Sum, Mean absolute change, Energy, Auto regressive, Skewness, Kurtosis
Frequency domain	Bands energy (0-255 Hz divided into 5 bands)

I. Multi-Camera Storage Area Occupancy Check

In the 1F inbound area, the ceiling height is 6 meters, allowing full-area coverage using ceiling-mounted cameras. However, for floors 2F to 5F, where the ceiling height is lower, ceiling-mounted cameras would require a large number of units, making the setup impractical. From a cost perspective, obliquely mounted overhead cameras are preferred. As a result, a single camera cannot fully capture the entire storage area due to occlusions, necessitating the integration of estimation results from multiple cameras.

In our work [30], we proposed a method to address this issue. Specifically, as illustrated in Fig. 14, we apply projective

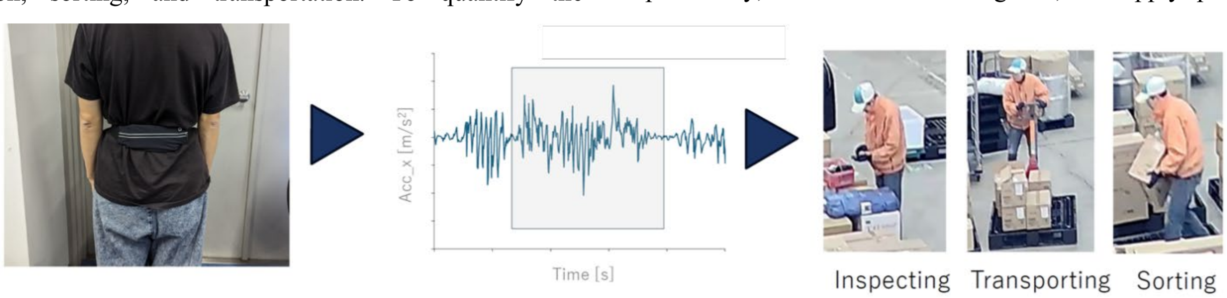

Fig.12. Task estimation using smartphone IMU sensors

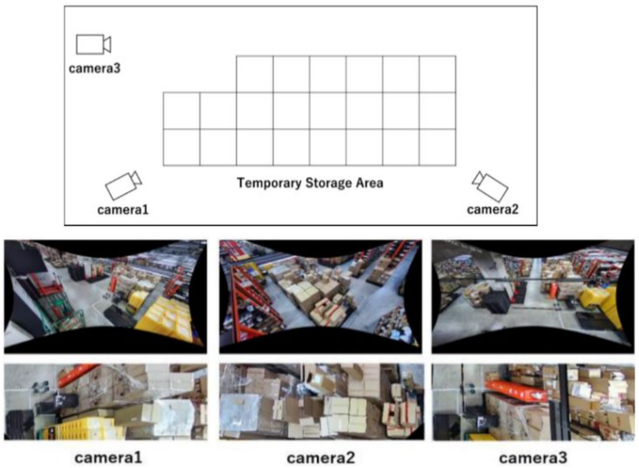

Fig.14. Camera images surrounding storage area.

transformation to convert images captured by cameras surrounding the put-away area into top-down views of the target area. Since each camera is affected by occlusions in different locations, we integrate the transformed images with varying weights to accurately estimate cargo space occupancy for each area.

IV. Black box optimization with quantum annealing

A Following the digitization of worker movement, the next step is to optimize worker shift schedules. Since conducting various trials in an actual warehouse would be prohibitively expensive, we developed a warehouse operations simulator. This simulator takes worker shifts, warehouse layout, and cargo information as inputs and simulates warehouse operations to output key performance metrics such as lead time for storage and total worker labor hours.

Building upon this simulator, we implement Factorization Machine Quantum Annealing (FMQA) [31], a black-box optimization technique. Factorization Machines (FM) are machine learning models capable of achieving high prediction accuracy for sparse input-output combinations, and the resulting mathematical model can be optimized using Quantum Annealing (QA).

Fig. 15 illustrates the black-box optimization process implemented in [32]. First, we generate a training dataset for the warehouse simulator by simulating operations with several initial parameter sets (worker shifts). Next, we train an FM model on this dataset. The learned mathematical model is then converted into a Quadratic Unconstrained Binary Optimization (QUBO) formulation, which is used for quantum optimization. The optimized worker shift schedule is re-evaluated using the simulator, generating additional training data to refine the FM model. This iterative process continues, progressively improving the optimization results. By applying FMQA to a combinatorial optimization problem with 21,300 possible configurations, we conducted an iterative optimization process using 500 initial data points and 100 iterations of FMQA. As a result, the proposed method successfully reduced the lead time by up to 37.4%, the residual cargo volume by up to 95.5%, and the total labor hours by up to 14.3%, demonstrating the effectiveness of the approach in optimizing warehouse operations.

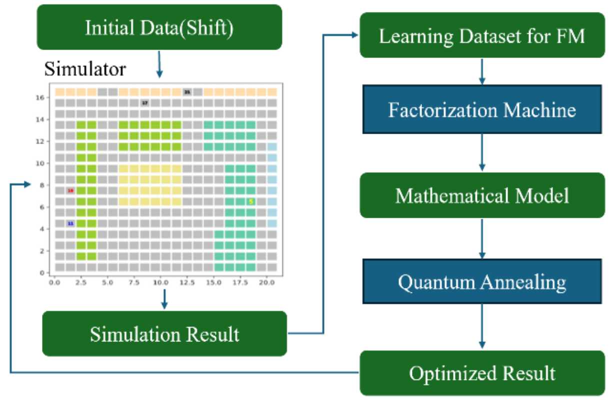

Fig.15. Blackbox optimization of warehouse using FMQA.

V. Visualization and Analysis Tool

To integrate the data obtained from the Camera Array Platform and Task Estimation, we developed a visualization and analysis tool, as illustrated in Fig. 16, enabling warehouse personnel to monitor and analyze operations effectively. This tool represents workers and cargo as three-dimensional objects, allowing users to observe warehouse conditions from any angle at any given time.

Furthermore, the framework provides functionalities for visualizing and analyzing worker tasks, cargo dwell times, and other operational insights. Continuous data analysis using this tool has progressively clarified various challenges within the warehouse, facilitating further optimization and operational improvements..

VI. Conclusion

This paper presents various initiatives we have undertaken to digitize and optimize large-scale logistics warehouses. The digitization of warehouses is a crucial step in achieving optimal logistics, and even the simple act of presenting the obtained data to workers through visualization tools has led to numerous insights. The current system is implemented using data stored in a database, but efforts are underway to enable real-time visualization and presentation of warehouse conditions. While this study focuses on a specific warehouse as a case example, the application of these methods to different warehouses in a short period requires the development of various generalization mechanisms.

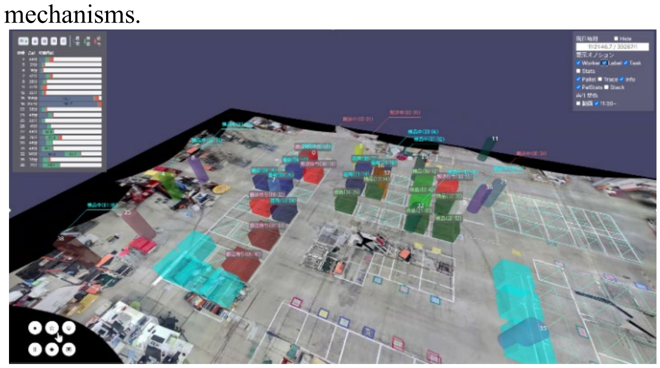
Fig.16. 3D Visualization and Analysis tool.

Acknowledgment

This work is partially supported by JSPS KAKENHI (JP22K18422), NEDO (JPNP23003), NICT (222C01, 22609) and CSTI SIP3 (JPJ012495).

References

[1] K. Aravindaraj, and P. Rajan Chinna, “A systematic literature review of integration of industry 4.0 and warehouse management to achieve Sustainable Development Goals (SDGs).” Cleaner Logistics and Supply Chain, Vol. 5, 100072, 2022. doi:10.1016/j.clscn.2022.100072.

[2] C. Rennie, R. Shome, K. E. Bekris, and A. F. De Souza, "A dataset for improved RGBD-based object detection and pose estimation for warehouse pick-and-place," IEEE Robotics and Automation Letters, vol. 1, no. 2, pp. 1179-1185, 2016, doi: 10.1109/LRA.2016.2532924.

[3] N. Correll et al., "Analysis and Observations From the First Amazon Picking Challenge," in IEEE Transactions on Automation Science and Engineering, vol. 15, no. 1, pp. 172-188, Jan. 2018, doi: 10.1109/TASE.2016.2600527.

[4] T. Li, B. Huang, C. Li, and M. Huang, “Application of convolution neural network object detection algorithm in logistics warehouse,” 7th International Symposium on Test Automation and Instrumentation (ISTAI 2018), 2018.

[5] J. Schyga, J. Hinckeldeyn, B. Bruss, C. Bamberger, and J. Kreutzfeldt, "Application-driven test and evaluation framework for indoor localization systems in warehouses," Eleventh International Conference on Indoor Positioning and Indoor Navigation(IPIN 2021), 2021.

[6] N. Yoshimura, J. Morales, T. Maekawa and T. Hara, "OpenPack: A large-scale dataset for recognizing packaging works in IoT-enabled logistic environments," 2024 IEEE International Conference on Pervasive Computing and Communications (PerCom), pp. 90-97, 2024. doi: 10.1109/PerCom59722.2024.10494448.

[7] T. Xie, and X. Yao, “Smart logistics warehouse moving-bbject tracking based on YOLOv5 and DeepSORT,” Appl. Sci. 2023, 13(17), 9895; doi:10.3390/app13179895.

[8] Ultralytics. “YOLOv5: Scalable and SOTA object detection models,” GitHub repository. Retrieved from https://github.com/ultralytics/yolov5

[9] N. Wojke, A. Bewley and D. Paulus, "Simple online and realtime tracking with a deep association metric," 2017 IEEE International Conference on Image Processing (ICIP), pp. 3645-3649, 2017. doi: 10.1109/ICIP.2017.8296962.

[10] Y. Asai et al., "Towards a real-time and energy-efficient edge AI camera architecture in mega warehouse environment," 2024 IEEE 3rd Real-Time and Intelligent Edge Computing Workshop (RAGE), Hong Kong, Hong Kong, pp. 1-6, 2024. doi: 10.1109/RAGE62451.2024.00010.

[11] Y. Mori, Y. Asai, K. Higashiura, S. Katayama, K. Urano, T. Yonezawa and N. Kawaguchi, "Efficient edge AI-based annotation and detection framework for logistics warehouses," IEEE Consumer Communications & Networking Conference (CCNC 2025), 2025.

[12] V. Usenko, N. Demmel and D. Cremers, "The double sphere camera model," 2018 International Conference on 3D Vision (3DV), Verona, Italy, 2018, pp. 552-560, doi: 10.1109/3DV.2018.00069.

[13] Leica Geosystem, BLK2Go,https://leica-geosystems.com/products/laser-scanners/autonomous-reality-capture/leica-blk2go-handheld-imaging-laser-scanner , accessed 2025/3/10.

[14] D. DeTone, T. Malisiewicz, and A. Rabinovich. "Superpoint: Self-supervised interest point detection and description," Proceedings of the IEEE conference on computer vision and pattern recognition workshops(CVPR2018). 2018.

[15] P. Lindenberger, P. E. Sarlin and M. Pollefeys, "Lightglue: Local feature matching at light speed." Proceedings of the IEEE/CVF International Conference on Computer Vision(ICCV2023). 2023.

[16] Y. Wu, A. Kirillov, F. Massa, W. Y. Lo and R. Girshick, “Detectron2,” https://github.com/facebookresearch/detectron2, 2019. (Accessed : 2025-03-01).

[17] K. Higashiura et al., "Semi-automated framework for digitalizing multi-product warehouses with large-scale camera arrays," 2024 IEEE International Conference on Pervasive Computing and Communications (PerCom), pp. 98-105, 2024. doi: 10.1109/PerCom59722.2024.10494498.

[18] Y. Mori, Y. Asai, K. Higashiura, S. Katayama, K. Urano, T. Yonezawa and N. Kawaguchi, "Efficient annotation method for object detection model creation for edge AI," Multimedia, Distributed, Cooperative, and Mobile Symposium (DICOMO2024), 2024, (in Japanese).

[19] Z. Teed, and J. Deng, " RAFT: Recurrent All-Pairs field transforms for optical flow, " ECCV 2020: 16th European Conference, 2020. doi: 10.1007/978-3-030-58536-5_24.

[20] X. Chen and K. He, “Exploring simple Siamese representation learning,” In 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 15745-15753, 2021.

[21] L. McInnes, J. Healy, N. Saul, and L. Grossberger., “UMAP: Uniform Manifold Approximation and Projection,” The Journal of Open Source Software, Vol. 3, No. 29, p. 861, 2018.

[22] K. Kano, Y. Mori, K. Higashiura, T. Hossain, S. Katayama, K. Urano, T. Yonezawa and N. Kawaguchi, "Composite image generation using labeled segments for pattern-rich dataset without unannotated target," Companion of the 2024 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 507-512, 2024. doi: 10.1145/3675094.3678447.

[23] K. Yokoyama, S. Katayama, K. Urano, T. Yonezawa and N. Kawaguchi, "Digitization and analysis framework for warehouse truck berth," 2023 Fourteenth International Conference on Mobile Computing and Ubiquitous Network (ICMU), Kyoto, Japan, pp. 1-4, 2023. doi: 10.23919/ICMU58504.2023.10412228.

[24] T. Meinhardt, A. Kirillov, L. Leal-Taixé, and C. Feichtenhofer, "TrackFormer: Multi-Object Tracking with Transformers," 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 8834-8844, doi: 10.1109/CVPR52688.2022.00864.

[25] Y. Zhang, P. Sun, Y. Jiang, D. Yu, F. Weng, Z. Yuan, P. Luo, W. Liu, and X. Wang, "Bytetrack: Multi-object tracking by associating every detection box," in Proc. Eur. Conf. Comput. Vis. (ECCV2022), 2022, pp. 1–21.

[26] N. Kawaguchi, J. Nozaki, T. Yoshida, K. Hiroi, T. Yonezawa, and K. Kaji, "End-to-end walking speed estimation method for smartphone PDR using DualCNN-LSTM," Proc. of The Tenth International Conference on Indoor Positioning and Indoor Navigation(IPIN2019), 2019.

[27] K. Kano, T. Yoshida, S. Katayama, K. Urano, T. Yonezawa and N. Kawaguchi, "Gait-robust heading estimation using horizontal acceleration for smartphone-based PDR," 2023 Proceedings of the Work-in-Progress Papers at the 13th International Conference on Indoor Positioning and Indoor Navigation (IPIN-WiP 2023), 2023.

[28] K. Kano, T. Yoshida, N. Hayashida, Y. Asai, H. Matsuyama, S. Katayama, K. Urano, T. Yonezawa and N. Kawaguchi, "Smartphone localization with solar-powered BLE beacons in warehouse," HCI International 2022, 2022.

[29] K. Watanabe, K. Kano, S. Katayama, K. Urano, T. Yonezawa and N. Kawaguchi, "Task recognition of workers using smartphones in logistics warehouses," The Joint Conference of Institutes of Electrical and Electronics Engineers, Tokai Section, 2024, (in Japanese).

[30] Y. Gushi, K. Kano, S. Katayama, K. Urano, T. Yonezawa and N. Kawaguchi, "Occupancy detection of storage space using occlusion avoidance with multi-camera," The Joint Conference of Institutes of Electrical and Electronics Engineers, Tokai Section, 2024, (in Japanese).

[31] K. Kitai, J. Guo, S. Ju, S. Tanaka, K. Tsuda, J. Shiomi, and R. Tamura, “Designing metamaterials with quantum annealing and factorization machines,” Physical Review Research, Vol. 2, No. 1, 013319, 2020.

[32] K. Takaki, K. Takigami, S. Katayama, K. Urano, T. Yonezawa and N. Kawaguchi, "Warehouse work optimization using quantum annealing and black-box optimization," UBI Research Group, Information Processing Society of Japan, vol. 2024-UBI-82, pp.1-8, 2024, (in Japanese).