Enhancing Warehouse Management with AI and Computer Vision: A Case Study in a Logistics Service Company

Nalinya Utamapongchai[1*], Sirawich Ngernsalung[2], Raveekiat Singhaphandu[3], and Warut Pannakkon[4*]

[1,2,3,4]School of Manufacturing Systems and Mechanical Engineering,
Sirindhorn International Institute of Technology, Thammasat University,
Pathum Thani, Thailand

[3]School of Knowledge Science, Japan Advanced Institute of Science and Technology,
Ishikawa, Japan

E-mail: nalinya.utm@gmail.com*, sirawich.nge@dome.tu.ac.th, raveekiat@gmail.com,
warut@siit.tu.ac.th

Received: June 21, 2024 / Revised: September 6, 2024 / Accepted: September 17, 2024

Abstract— **In the evolving landscape of warehouse management in Industry 4.0, this paper explores the convergence of Artificial Intelligence (AI) and Computer Vision (CV) for inventory tracking and stock registration. Conducted in collaboration between SIIT and KNS, a logistics service company specializing in warehousing, the study introduces a framework that optimizes image capture conditions through real-time analysis of gyroscope values, distinguishing mobile phone movement from stationary states. Additionally, an object detection model using the YOLOv8 algorithm achieves 83% accuracy in label detection and 75% in box detection within a curated dataset. The research highlights the successful development of the phone motion detection model and Optical Character Recognition (OCR) integration. This framework promises to advance warehouse management systems by addressing current limitations with a comprehensive, efficient, and user-friendly solution.**

Index Terms— **Digital Transformation, Artificial Intelligence, Computer Vision, Inventory Tracking**

I. Introduction

In the present time, it is a transformative era, where the world is rapidly progressing towards a new age characterized by innovative and technological revolution [1]. This age is also defined by the convergence of numerous advancements, essential among them being the development of Industry 4.0 and the extraordinary growth of Artificial Intelligence (AI) [1], [2].

The rapid growth of AI does not only enable more applications in various fields, but it also enhances the capabilities of most of its features, including computer vision technology [1], [2]. Computer vision is referred to as a field of AI that enables computers and systems to derive meaningful information from digital images, videos, and other visual inputs [3], [4]. Before the AI era, computer vision primarily relied on handcrafted algorithms, which were time-consuming and low in capacity and accuracy. However, as deep learning becomes predominant, it has now become the core part of computer vision [3], [4]. This significant change allows the algorithms to handle more complex and diverse scenes with precise results [1], [3].

As a result, modern computer vision technology is utilized in numerous fields. Plus, with the development of big data and the Internet of Things (IoT), a powerful, well-balanced integrated management system is established [2], [3], [5]. Some of the fields gaining the most benefit from this system are warehousing and inventory management [2], [6].

However, most warehouse management technologies offer automation, visualisation, and inventory tracking, suitable for large-scale systems. Challenges include inflexibility to system changes, high initial costs, and complex, immobile control programs [2], [6], [7]. Barcode scanners, the primary outgoing product check tool, have limitatio ns, scanning one product at a time and causing operator fatigue [6], [8]. A significant gap exists in tracking stocks moving through the system without proper registration, hindering analysis of stock flow duration in the warehouse.

Thus, with the engagement of AI along with Optical Character Recognition (OCR) technology in computer vision, this study aims to achieve the following objectives: (1) Develop a phone motion detection model that optimizes image capture conditions through real-time analysis of gyroscope values; (2) Apply the knowledge of computer vision and OCR to develop box and label detection models for a stock registration/record system; (3) Evaluate the performance of the developed models. The rest of the paper is organized as follows: Part II - Related Work, including literature review and relevant theory, Part III - Methodology, including project framework and performance evaluation, Part IV - Case Study, including the procedures, Part V - Result and Discussion, and Part VI - Conclusion and Future Work.

II. Related Work

In the domain of Artificial Intelligence (AI), Computer Vision (CV) plays a pivotal role by enabling machines to comprehend visual information akin to human vision [4]. This extends from image and video analysis to intricate tasks such as object recognition, tracking, and scene understanding [7], [9], [10]. The integration of CV algorithms and Optical Character Recognition (OCR) techniques augments its capabilities, allowing localization of regions of interest and text recognition [11], [12].

In the context of Industry 4.0, a transformative shift towards a connected, automated, and data-driven approach is witnessed, with key components including the Internet of Things (IoT), Big Data and Analytics, AI, and Machine Learning (ML) [4], [5], [10]. The emergence of smart Warehouse Management Systems (WMS) exemplifies the interconnected use of AI, CV, and IoT to optimize warehouse processes, aiming for improved accuracy and efficiency [2], [6]. This holistic approach spans stock planning, product placement, order picking, transport, and tracking within the warehouse workflow, presenting opportunities for optimization in each process.

In the retail sector, CV's rapid growth has led to a focus on product image recognition systems [9]. This framework encompasses image capture, pre-processing, feature extraction, feature classification, and recognition output, offering a systematic approach to image-based recognition.

The emphasis on high-quality input images is critical for the success of CV applications, underlining the importance of meticulous image management from the initial capture phase [13]. Simultaneously, the investigation into the utility of built-in tri-axial accelerometer and gyroscope sensors in smartphones has been conducted across various scenarios involving human motion [14], [15]. This exploration extends to diverse activities, including the recognition of hand gestures, and extends further by incorporating additional wrist-worn sensor attachments [16]. The seamless integration of these technologies contributes to a comprehensive understanding of motion, enabling enhanced applications in image recognition and other CV-related tasks.

Additionally, the YOLO (You Only Look Once) algorithm is scrutinized for real-time object detection, assessing its characteristics and performance. Various versions of YOLO have been developed, incorporating innovative ideas and techniques, offering researchers new directions for addressing challenges in object detection [17]-[19].

In conclusion, the comprehensive review of related work demonstrates the dynamic evolution of intelligent systems, highlighting the integration of AI, CV, and IoT in Industry 4.0. The utilities of gyroscopes and accelerometers in smartphones have the potential to be applied as key facilitators in the image acquisition stage, effectively reducing barriers and skill requirements for operators. These findings collectively contribute to the ongoing exploration of advanced technologies for enhanced image recognition and object detection in the research presented in this paper.

III. Methodology

This part consists of two main sections: A discussion of the generalized concept of the proposed framework and a performance evaluation. Fig. 1 illustrates the framework of the proposed system for double-checking outbound processes.

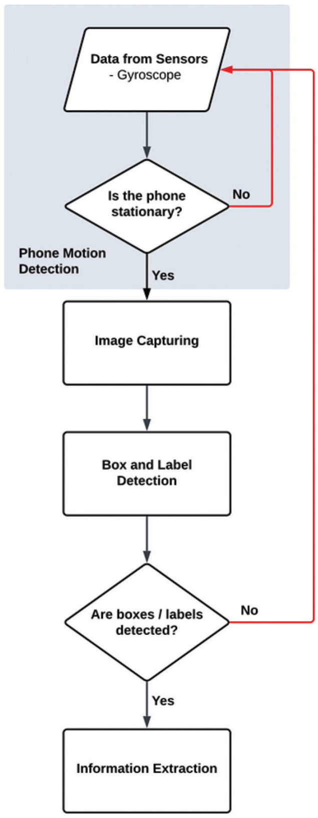

Fig. 1. Framework of the proposed system

A. Proposed Framework

The proposed methodology starts with a foundational hypothesis suggesting that images captured under stationary conditions with a mobile phone inherently have good quality, leading to successful OCR outcomes. To validate this hypothesis, the model was developed using Android Studio with the capability to identify whether the phone is stationary or in motion, which is crucial in determining image quality for OCR purposes. After that, the captured images undergo box and label detection models scoping down the area of interest before passing on to the information extraction step by the OCR process as illustrated in Fig. 2.

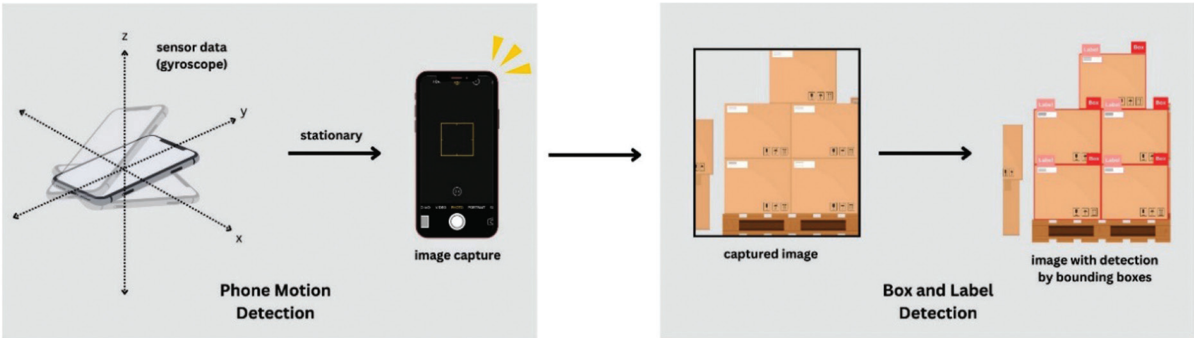

Fig. 2. System function

1) Phone Motion Detection

In practical terms, the application employs a comparative analysis of gyroscope values against predetermined threshold values. If the gyroscope values fall within the defined limits across all axes, the phone is considered stationary and ready for image capture. Conversely, if one or more gyroscope values surpass the established thresholds, the application identifies the phone as in motion.

In contrast to the gyroscope's comparative analysis, the application will not utilize accelerometer sensor values for motion detection. This decision is based on the consistent influence of gravitational force (g) on the phone, yielding sensor values around g (9.81 m/s^2) or negative g (-9.81 m/s^2), particularly influenced by the device's orientation.

During the phase of phone movement, the application should consistently exhibit a "NOT READY" status. It is anticipated that the readiness status will shift to "READY" when the phone comes to a stop and will persist in this status during stationary phases.

2) Box and Label Detection

The data collection phase began with the use of a mobile phone camera to capture images of boxes and labels from various angles and distances. This approach aimed to provide a diverse representation of real-world scenarios, allowing the model to extract relevant features and patterns essential for robust performance across different situations.

After the completion of the data collection phase, Roboflow played a pivotal role in streamlining the workflow. Its annotation tools facilitated the precise annotation of collected images with bounding boxes for both boxes and labels, ensuring accurate object localization. Moreover, Roboflow's augmentation techniques, including shear, blur, and noise, expanded the dataset and enhanced its diversity.

Leveraging Roboflow for image pre-processing, YOLOv8 effectively processed the augmented data to accurately detect and classify objects within the images. Capitalizing on YOLOv8's deep neural network architecture, the model demonstrated the capability to simultaneously detect multiple objects, including boxes and labels.

B. Performance Evaluation

The evaluations are required for two main processes of the framework, phone motion detection and box and label detection.

The evaluation of phone motion detection focuses on the application's ability to dynamically adapt its status indicators in response to real-time phone movements, ensuring an accurate depiction of readiness during stationary periods and non-readiness during mobile phases. This is achieved through the intricate processing of sensor data and threshold comparisons, which distinguish between stationary and moving states. With the important role of threshold value resulting in the sensitivity of motion detection, the value adjustment allows the application to suitably perform in the simulated scanning task, replicating the real-world use.

The evaluation of the box and label detection models involves a comprehensive examination of their performance in accurately identifying and categorizing boxes and labels within a predefined dataset.

The performance metrics are considered, with a specific emphasis on scrutinizing a confusion matrix. This matrix serves as a pivotal analytical tool, offering a nuanced understanding of the model's classification errors and providing valuable insights into specific challenges and areas for improvement, particularly in the domain of multi-label classification. The essential elements within the confusion matrix, encompassing True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN), provide insights into the model's strengths and weaknesses, especially in box and label classification scenarios, highlighting specific areas and allowing for targeted enhancements to the model's predictive capabilities.

IV. Case Study

In our case study, the proposed framework is applied to a real-world application to develop a system implemented in two main sections: phone motion detection and box and label detection.

A. Phone Motion Detection

In adherence to the methodology outlined earlier, our application conducts a real-time comparative analysis of gyroscope sensor values against a predefined threshold to discern the stationary or moving status of the mobile phone. During the tuning process, through trial-and-error experiments, this threshold is set at 0.2, taking into account specific characteristics of the mobile device and factors related to human interaction. This calibrated value strikes a balance, ensuring the phone remains stable enough for high-quality image capture while remaining manageable for users to stay within the specified limits.

As the application currently exists in a prototype stage, the image-capturing process is not activated immediately. Once the system confirms the phone's stationary state, the user interface will prominently display the status "READY" accompanied by a reassuring green box. This visual cue signals that the phone is primed for image capture. Conversely, if the gyroscope values exceed the threshold, the application communicates the status "NOT READY" along with a conspicuous red square, signaling a temporary pause in the image-capturing process. This streamlined feedback mechanism aligns with the application's intent to enhance user experience and maintain control over imaging conditions.

B. Box and Label Detection

1) Data Collection

In the objective of developing a robust object detection and classification model applicable in practical scenarios, a comprehensive dataset was collected from KNS Logistics Service Company Limited, a company offering extensive services in warehouse management, storage, and distribution of products. The dataset comprises images and videos capturing diverse dimensions and characteristics of boxes and labels within the warehouse and dock stations. A mobile phone camera was utilized during the data collection process to replicate the anticipated implementation method.

Each product was captured from multiple angles, including various distances to encompass a range of perspectives. The decision to incorporate both long-distance and close-up images was to ensure the clarity of information captured for both boxes and labels. The intentional inclusion of multiple perspectives enhances the model's adaptability to real-world scenarios, where products may be encountered at different distances. Sample images captured are illustrated in Fig. 3 to Fig. 5.

Fig. 3. Sample of the captured image of boxes

Fig. 4. Sample of the captured image of labels

Fig. 5. Sample of captured image

To enrich the dataset further, videos were recorded, showcasing individual labels or groups of labels continuously. This deliberate approach aimed to capture temporal variations and nuances in label presentation, contributing to a more comprehensive understanding of the visual context. In total, 140 images and video snapshots were initially prepared, and through special augmentation techniques provided by Roboflow, the dataset was expanded to a final count of 340 samples.

2) Model Development

Since the automatic image-capturing algorithm has been waiting for forthcoming development, the manually captured images have been used as the imitated datasets to ensure the performance of the detection model. The collected images were processed and prepared through Roboflow before being passed onto the YOLOv8 model for training by steps as follows:

• *Image Annotation*: Roboflow's annotation tools were utilized with 140 uploaded images containing objects to be detected and nulled images with bounding boxes, ensuring accurate annotations for boxes and labels.

• *Train/Test Split*: To contribute to the accuracy and reliability of the model, the dataset was randomly split into training, validating, and testing sets with the set of ratios (i.e., 70%:20%:10%).

• *Pre-Processing*: All images in the dataset underwent image transformations with the aim of reducing training time and enhancing performance. This was achieved through the application of the auto-orient and resize functions.

• *Augmentation*: New variations and the increase of training images were performed by augmentation application. The additional images for the training set of the model are generated from the existing in augmented versions with shear, blur, and noise effects, resulting in 340 total images.

• *Model Training*: The annotated and augmented dataset underwent conversion to a format compatible with YOLOv8. Subsequently, it was subjected to the training process, aiming to minimize the disparity between predicted and actual outputs.

V. Result

The mobile application prototype was rigorously evaluated for its performance, particularly focusing on its ability to correctly identify the phone's motion status during the experiments. The application demonstrated excellent performance, with no false results recorded throughout the testing phase. Specifically, as shown in Fig. 6, when the mobile phone was in motion, the application reliably indicated a 'NOT READY' status. Conversely, when the phone was stationary, it swiftly transitioned to the 'READY' status, maintaining this state effectively without requiring excessive stabilization efforts.

Fig. 6. The mobile application's User Interface (UI) displaying the current status

The box and label detection models were tested using a series of input images, with the results visualized in figures (Fig. 7 to Fig. 9). The detection models achieved a 75% accuracy rate in identifying boxes and an 83% success rate in detecting labels, as demonstrated by the confusion matrix in Fig. 10. However, the confusion matrix also highlighted areas of concern: the box detection model showed a 43% false positive rate, and the label detection model showed a 57% false positive rate. A possible cause of these false positives is that the training image sets contain boxes and labels at various distances, which may have caused the model to lack a three-dimensional understanding, leading it to mistakenly identify non-relevant objects as the object of interest. Additionally, some boxes have multiple colors and patterns, which can further confuse the model when distinguishing the object of interest.

Fig. 7. Result of multi-label detection for boxes

Fig. 8. Result of multi-label detection for labels

Fig. 9. Result of multi-label detection for boxes and labels

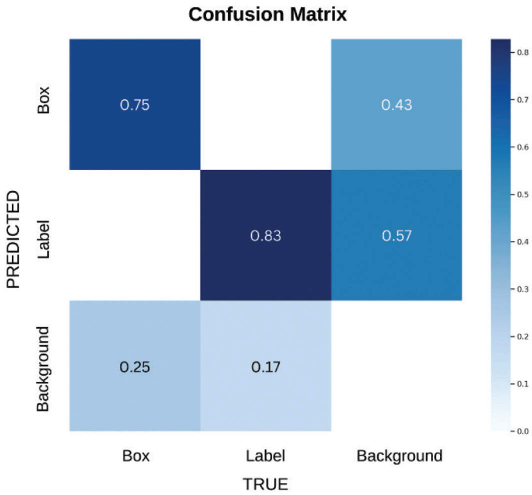

Fig. 10. Confusion matrix of multi-label detection

An example of a false positive case is shown in Fig. 11 and Fig. 12. Fig. 11 illustrates the human annotation (ground truth), while Fig. 12 displays the model's prediction. It appears that the model mistook the white rectangle on the top left corner as the label of interest, highlighting a specific scenario where the model's limitations are evident.

Fig. 11. Example of human annotation (ground truth)

Fig. 12. Example of false positive prediction

VI. Discussion

The sensitivity of the motion detection feature in the mobile application was found to be finely tuned, balancing the need to distinguish between minor, natural hand movements and significant motion. The selected threshold values were set to accommodate small shakes or sways that naturally occur due to hand fatigue, ensuring that these do not falsely trigger motion detection. This careful calibration minimizes errors and enhances the accuracy of the application's ability to detect when the phone is truly stationary versus in motion. Despite the success of the current motion detection capabilities, the application currently only displays the status ('READY' or 'NOT READY') without further action. Future developments should focus on integrating this motion detection with an automatic image-capturing process that activates when the device is in the 'READY' state, enabling continuous image capture and real-time data processing.

The results suggest that the box and label detection models may be experiencing overfitting, as indicated by the high false positive rates. Overfitting occurs when the model learns specific features from the training data that do not generalize well to new, unseen data, potentially mistaking noise for actual patterns. On the other hand, the false negative rates—25% for boxes and 17% for labels—point to underfitting, where the model fails to capture important details, resulting in missed detections. A critical observation is that the current dataset is more suited to label detection due to the close-up nature of the images, which limits the visibility of entire boxes. To improve the robustness and generalizability of the models, expanding the dataset to include a more diverse range of images is necessary. This expansion will likely reduce both overfitting and underfitting, enhancing the overall performance of the detection models.

VII. Conclusion and Future Work

The culmination of this study has resulted in the successful development of two integral components: The phone motion detection and the box and label detection models. The phone motion detection model adeptly utilizes real-time gyroscope values to discern the dynamic state of the phone, effectively distinguishing between periods of movement and stationary conditions. The performance evaluation, conducted through a simulated scanning process, underscores the application's excellence by consistently and accurately indicating the mobile phone's status.

In parallel, the box and label detection models are powered by a meticulously curated dataset comprising 140 images and videos of actual products collected during a company visit. Augmented and trained using the YOLOv8 model through Rob flow, the model demonstrates commendable performance metrics.

Notably, label detection achieves an impressive 83% accuracy, considering the limited training samples, while box detection lags with 75% accuracy. This disparity can be attributed to the nature of the collected samples, which favoured label detection due to their proximity to the boxes, making the latter challenging to discern clearly.

While both components have met expectations and demonstrated exceptional individual performance, acknowledging the need for revision and enhancement is crucial. Further refinement is essential before amalgamating the mobile application prototype and the object detection and classification model into a comprehensive and polished version of the application. This research lays a solid foundation for future iterations, emphasizing the continuous pursuit of excellence in mobile application development and object detection technology.

In considering future work, several key recommendations emerge to enhance the overall functionality and performance of both the mobile application and the object detection and classification model. For the phone motion detection model, an imperative focus lies in developing a streamlined mechanism to access the mobile phone's camera. This feature is crucial to enable the application to initiate image capture during periods when the status is deemed ready. Additionally, there is a clear opportunity to refine and elevate the User Interface (UI) to ensure a simpler, more professional appearance, incorporating all essential information seamlessly for user clarity and engagement.

Furthermore, an experiment on the gyroscope's threshold values is warranted to ascertain the optimal settings, striking a balance between minimizing false results and ensuring a user-friendly configuration.

For the box and label detection models, it is recommended to conduct a prospective visit to the company for the collection of more samples, specifically long-distance shots, to enhance the performance of box detection. Continual training iterations for the model should be pursued to ensure sustained improvement in accuracy and overall performance. Additionally, integrating OCR into the model is a promising avenue to extract serial number information from detected labels, contributing valuable data for subsequent processing. The consideration of using PPOCR should be processed in this further step as its ease of use, lightweight design, and minimal computational requirements. It supports multiple languages, handles text in various orientations and backgrounds, and is backed by good documentation and community support, simplifying integration and troubleshooting. Incorporating phone motion detection with the box and label detection models, future efforts should concentrate on establishing a consistent mechanism to pass captured images to the model for further processing. Implementing a robust storage solution, such as a database, will be essential for efficiently storing the extracted information. Furthermore, efforts can be directed toward linking the extracted information with the sales order database to facilitate a comprehensive double-check process, thereby enhancing the overall accuracy and reliability of the system.

Acknowledgments

The authors wish to express their sincere appreciation to KNS Logistics Service Company Limited for graciously providing invaluable information and industry insights that greatly supported this study.

References

[1] R. S. Peres, X. Jia, J. Lee et al., "Industrial Artificial Intelligence in Industry 4.0-Systematic Review, Challenges and Outlook," *IEEE Access*, vol. 8, pp. 220121-220139, Dec. 2020.

[2] Ö. Albayrak Ünal, B. Erkayman, and B. Usanmaz, "Applications of Artificial Intelligence in Inventory Management: A Systematic Review of the Literature," *Arch. Compute. Methods Eng.*, vol. 30, no. 4, pp. 2605-2625, Feb. 2023.

[3] M. D. Capua, A. Ciaramella, and A. De Prisco, "Machine Learning and Computer Vision for the Automation of Processes in Advanced Logistics: The Integrated Logistic Platform (ILP) 4.0," *Procedia Comput. Sci, 217*(9), vol. 217, no. 9, pp. 326-338. doi.org/10.1016/j.procs.2022.12.228

[4] M. Javaid, A. Haleem, R. P. Singh et al., "Exploring Impact and Features of Machine Vision for Progressive Industry 4.0 Culture," *Sensors Int*, vol. 3, no. 5, p. 1000132, Nov. 2021. p. 100132. doi.org/10.1016/j.sintl.2021.100132

[5] Y. Song, F. R. Yu, L. Zhou et al., "Applications of the Internet of Things (IoT) in Smart Logistics: A Comprehensive Survey," *IEEE Internet Things Journal*, vol. 8, no. 6, pp. 4250-4274, Mar. 2021. doi.org/10.1109/jiot.2020.3034385

[6] E. Zunic, S. Delalic, K. Hodzic et al., "Smart Warehouse Management System Concept with Implementation," in *Proc. 2018 14th Symposium on Neural Networks and Applications (NEUREL)*, 2018, pp. 1-5. doi.org/10.1109/neurel.2018.8587004

[7] A. Vukicevic, M. Mladineo, N. Banduka et al., "A Smart Warehouse 4.0 Approach for the Pallet Management Using Machine Vision and Internet of Things (IoT): A Real Industrial Case Study," *Adv. Prod. Eng. Manag*, vol. 16, no. 3, pp. 297-306, Sep. 2021. doi.org/10.14743/apem2021.3.401-

[8] R. Wudhikarn, P. Charoenkwan, and K. Malang, "Deep Learning in Barcode Recognition: A Systematic Literature Review," *IEEE Access*, vol. 10, pp. 8049-8072, Jan. 2022. doi.org/10.1109/access.2022.3143033

[9] Y. Wei, S. Tran, S. Xu et al., "Deep Learning for Retail Product Recognition: Challenges and Techniques," *Comput. Intell. Neurosci*, vol. 2020, pp. 1-23, Nov. 2020.

[10] G. Monteiro, L. Camelo, G. Aauino et al., "A Comprehensive Framework for Industrial Sticker Information Recognition Using Advanced OCR and Object Detection Techniques," *Appl. Sci.*, vol. 13, no. 12, p. 7320, May. 2023. doi.org/10.3390/app13127320

[11] S. Long, X. He, and C. Yao, "Scene Text Detection and Recognition: The Deep Learning Era," *Int. J. Comput. Vis.*, vol. 129, no. 1, pp. 161-184, Aug. 2020. doi.org/10.1007/s11263-020-01369-0

[12] X. Chen, L. Jin, Y. Zhu et al., "Text Recognition in the Wild," *ACM Comput. Surv.*, vol. 54, no. 2, pp. 1-35, Dec. 2021. doi.org/10.1145/3440756

[13] X. Yang, F. Sang, T. Wang et al., "Research on the Influence of Camera Velocity on Image Blur and a Method to Improve Object Detection Precision," in *Proc. The 21 International Conference on Cyber-Physical Social Intelligence* (ICCSI), 2021, pp. 1-6. doi.org/10.1109/iccsi53130.2021.9736224

[14] I. M. Pires, N. M. García, E. Zdravevski et al., "Activities of Daily Living with Motion: A Dataset with Accelerometer, Magnetometer and Gyroscope Data from Mobile Devices," *Data Brief*, vol. 33, p. 106628, Dec. 2020. doi.org/10.1016/j.dib.2020.106628

[15] A. Jain and V. Kanhangad, "Human Activity Classification in Smartphones using Accelerometer and Gyroscope Sensors," *IEEE Sensors Journal*, vol. 18, no. 3, pp. 1169-1177, 2018. doi.org/10.1109/jsen.2017.2782492

[16] M. Shoaib, S. Bosch, Ö. D. İncel et al., "Complex Human Activity Recognition Using Smartphone and Wrist-Worn Motion Sensors," *Sensors*, vol. 16, no. 4, p. 426, Feb. 2016. doi.org/10.3390/s16040426

[17] P. Jiang, D. Ergu, F. Liu et al., "A Review of Yolo Algorithm Developments," *Procedia Comput. Sci.*, vol. 199, pp. 1066-1073, Feb. 2022.

[18] T. Diwan, G. Anirudh, and J. V. Tembhurne, "Object Detection Using YOLO: Challenges, Architectural Successors, Datasets and Applications," *Multimedia Tools Appl.*, vol. 82, no. 6, pp. 9243-9275, Aug. 2022. doi.org/10.1007/s11042-022-13644-y

[19] C. H. Kang and S. Y. Kim, "Real-Time Object Detection and Segmentation Technology: An Analysis of the YOLO Algorithm," *JMST Advances*, vol. 5, no. 2-3, pp. 69-76, Sep. 2023. doi.org/10.1007/s42791-023-00049-7

Nalinya Utamapongchai is currently a fourth-year student, pursuing a Bachelor of Engineering in Industrial Engineering and Logistics Systems at Sirindhorn International Institute of Technology (SIIT), Thammasat University, Thailand. Her research interests encompass supply chain management, digital transformation, and data analysis.

Sirawich Ngernsalung is currently a fourth-year student in the Bachelor of Engineering program in Industrial Engineering and Logistics Systems at Sirindhorn International Institute of Technology (SIIT), Thammasat University, Thailand. Additionally, he is enrolled as an external student in the master's program of Logistics and Supply Chain Systems Engineering at SIIT. His research interests include lean automation, digital transformation, and data analytics.

Raveekiat Singhaphandu received a B.S. degree in computer science from the Sirindhorn International Institute of Technology (SIIT), Thammasat University, Thailand, in 2014, and an M.S. degree in informatics from the Technical University of Munich, Germany, in 2017. He is currently pursuing a Ph.D. degree with SIIT, Thammasat University, and the Japan Advanced Institute of Science and Technology, Japan. His research interests include machine learning, computer vision, and applied software engineering in industrial engineering applications, Science and Technology, Japan. His research interests include machine learning, computer vision, and applied software engineering in industrial engineering applications.

Warut Pannakkong is an Associate Professor of the School of Manufacturing Systems and Mechanical Engineering (MSME), Sirindhorn International Institute of Technology (SIIT), Thammasat University, Thailand. He received his B.Engr. in Industrial Engineering and M. Eng. in Logistics and Supply Chain Systems Engineering both from SIIT in 2010 and 2014, respectively. He earned his Ph.D. in Knowledge Science from Japan Advanced Institute of Science and Technology (JAIST) in 2017. His research interests include artificial intelligence for industry, time series forecasting, data mining, machine learning, computer vision in industrial applications, discrete event system simulation and optimization, production planning, and logistics and supply chain management.