Smart Warehouse Safety: Computer Vision for Forklift Driver Monitoring in Warehouse Setting

Apirak Sang-ngenchai
Kanazawa Institute of Technology
sapirak@neptune.kanazawa-it.ac.jp

Minoru Nakazawa
Kanazawa Institute of Technology
Nakazawa@infor.kanazawa-it.ac.jp

Abstract

Safety for forklifts and warehouse workers must be prioritized. Our study introduces a computer vision-based solution for promptly recognizing unsafe behaviors in real-time using deep learning and pose estimation. The primary goal is to identify unsafe actions of forklift drivers, enhance workplace safety, and help managers spot dangerous trends or patterns early on. The research will integrate Convolutional Neural Networks and Pose Estimation. Specifically, we will use the "You Only Look Once" architecture version 8 of the CNN model for object detection to identify forklift drivers. Subsequently, the YOLO version 8 pose will detect and estimate forklift driver body joints. Our findings indicate that this approach achieves a detection accuracy rate of over 91 percent and an f1 score of the pose estimation model exceeding 0.96. Importantly, the potential of our solution is not limited to forklift safety in warehouses. Future work will focus on adapting the model to different warehouse settings and other high-risk industries, demonstrating our research's versatility and wide-ranging applicability.

Keywords

computer vision, object detection, pose estimation, forklift safety monitoring

I. Introduction

Forklifts are motorized machines used for lifting, unloading, and moving items. Due to their versatility and ease of use, they are commonly used in warehouses, manufacturing plants, and construction sites. However, in the United States, OSHA [1] reports 34,900 serious forklift accidents and 61,800 non-serious injuries related to forklift accidents annually. The Japan International Center for Occupational Safety and Health (JICOSH) [2] also reports that nearly 2,000 workers died annually from industrial accidents in the 2000s. Integrating technology in forklifts and warehouses is crucial for improving safety and reducing accidents caused by human error. Image detection and pose estimation systems can accurately track operators and help mitigate risky situations, providing a sense of reassurance and security. Implementing a monitoring system allows managers to identify and address unsafe trends in forklift operations.

In this study, the main benefit of the proposed framework is to enhance the safety and efficiency of forklift operations within a warehouse setting. The contributions of the work can be summarized as the following topics:

1. Gather and collect data from the existing CCTV camera to train an Object Detection Model and Pose Estimation Model for warehouse settings.
2. Develop a Pose Estimation Model to identify the unsafe behavior of the forklift drivers.
3. Evaluate the custom model using k-fold cross-validation techniques and classification reports. The proposed method obtains over 91 percent detection performance and 0.96 F1-Score for pose estimation.

II. Proposed Framework

The proposed framework utilizes computer vision to monitor safety and detect unsafe behavior among forklift drivers. It comprises three integrated modules: Object Detection, Pose Estimation, and Monitoring System. These modules operate within a central server, processing and storing data in a local database. The system is designed to notify workers when an unsafe action has occurred. It leverages existing warehouse cameras to stream real-time images to the server for processing by the Object Detection and Pose Estimation modules.

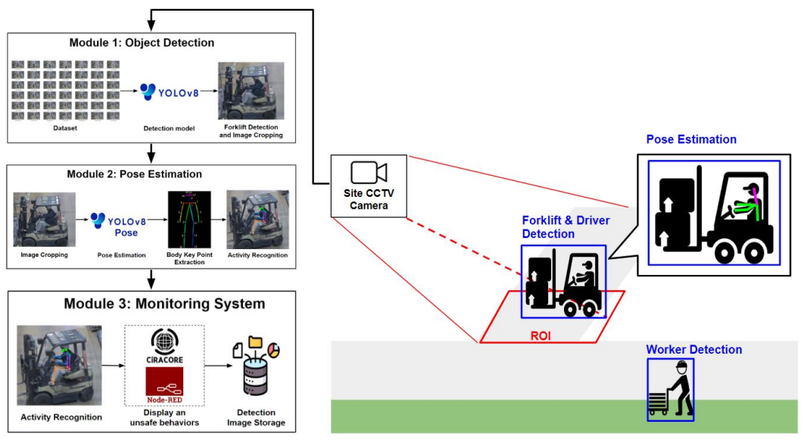

Fig. 1. The Proposed Framework Diagram

Module 1: Object Detection-based YOLOv8

Our first module focuses on creating a deep-learning model to classify workers, forklift trucks, and drivers using the YOLOv8 Architecture. We have a dataset of 5,417 images, segmented into Train (80%), Validation (10%), and Test (10%) Datasets. We use the GTGenerator tool from CiRA Core IDE for automatic labeling and adjusting parameters to train an accurate and efficient object detection model.

Module 2: Pose Estimation-based YOLOv8 pose

We have implemented the YOLOv8 pose model in CiRA Core IDE to extract the forklift driver's key points. The key points have been stored as CSV files for pose model training. The dataset has 1,340 rows, each containing 17 body key points, resulting in 34 coordinate values (x and y axes for each human body key point).

Module 3: Monitoring System

To monitor the unsafe behavior, we utilized the CiRA Core IDE to display detected information on a computer monitor. Red indicates unsafe activity, and green indicates safe activity. When unsafe activities are detected, data is captured in images and saved as JPG files in local storage. Additionally, we developed a Node-RED UI to showcase real-time streaming camera results, detection outputs, and indicators of unsafe activity. The monitoring of CiRA Core Workspace and Node-RED UI, as illustrated in Figures 2 and 3.

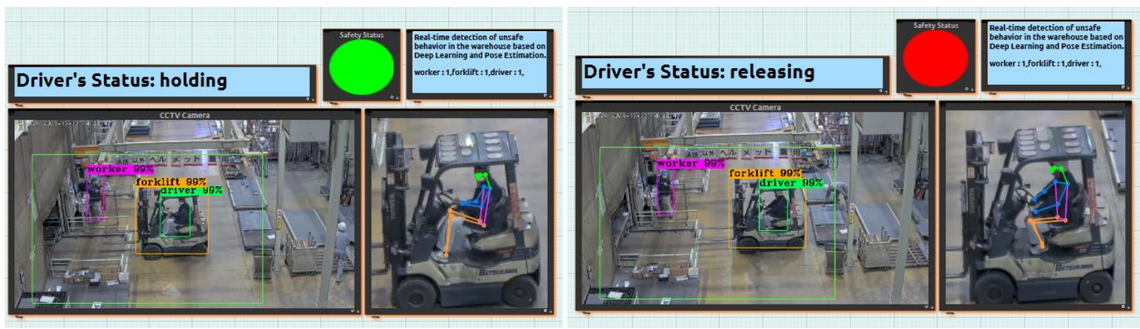

Fig. 2. CiRA Core Monitoring.

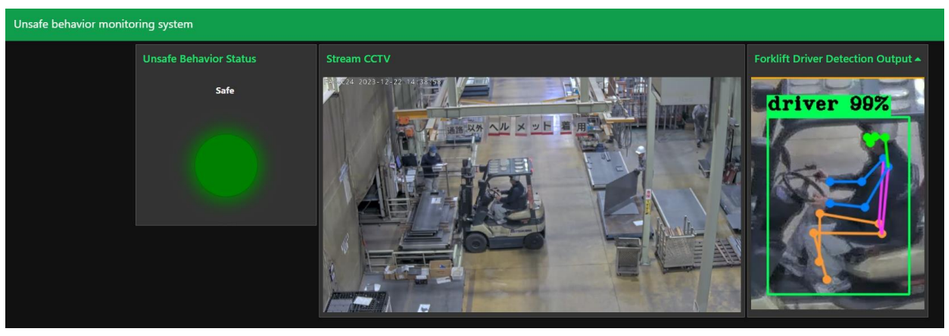

Fig. 3. Node-RED UI Monitoring.

III. Experiments and Results

The system was successfully trained and evaluated on a Windows PC with 16 GB of RAM, a 64-bit Operating System, an Intel(R) Core(TM) i7-12700H CPU, and an Nvidia GeForce RTX 3070 8 GB graphics processing unit (GPU). The upcoming section will discuss the results of the object detection model evaluation and pose estimation model evaluation.

A. Object Detection Model Evaluation

In our Object Detection Model, we compared our custom YOLOv8 model with state-of-the-art results. Using the K-Fold Cross-Validation method, we achieved 99.76% accuracy in the 5th K-Fold dataset. The average precision, recall, and accuracy are 0.913, 0.975, and 97.76%, respectively.

Table I. Object Detection Cross Validation (K=5)

	x_1	x_2	x_3	x_4	x_5	E
Precision	0.931	0.926	0.914	0.876	0.921	0.913
Recall	0.978	0.939	0.991	0.982	0.984	0.975
Accuracy%	99.30	91.45	98.84	99.53	99.76	97.76

B. Pose Estimation Model Evaluation

The report contains the classification results from our five models. All five pose estimation models achieved F1 scores between 0.96 and 0.99, with Architecture 2 having the highest testing accuracy and Architecture 4 the lowest. Prioritizing the F1 score over accuracy is crucial, especially in unbalanced problems.

Table II. Pose Estimation Architecture Classification Report

	Epoch	Train Accuracy %	Test Accuracy %	F1 Score
Architecture 1	26	0.9942	0.9845	0.98
Architecture 2	50	0.9968	0.9896	0.99
Architecture 3	17	0.9974	0.9689	0.97
Architecture 4	70	0.9948	0.9592	0.96
Architecture 5	170	0.9896	0.9796	0.98

IV. Conclusion

The paper presents a framework for identifying unsafe behaviors in forklift drivers using existing CCTV cameras. It utilizes object and human pose detection technologies and has demonstrated over 91% accuracy in detecting forklifts, drivers, and workers. Architecture 2 achieved the highest F1 score for Pose Estimation among the architectures tested. This framework's deployment necessitates investment and management approval for training and testing.

References
[1] Occupational Safety and Health Administration. (2023). Retrieved February 3, 2024, from https://www.osha.gov/ords/imis/accidentsearch.search?acc_keyword=forklift.
[2] Japan International Center for Occupational Safety and Health (JICOSH). Accident Cases. Retrieved February 3, 2024, from https://www.jisha.or.jp/english/information/accident_cases.html.

Acknowledgment
The authors express their gratitude to Betsukawa Corporation for providing the dataset and support. Additionally, they thank the College of Advanced Manufacturing Innovation at King Mongkut's Institute of Technology, Ladkrabang, Thailand, for providing the essential deep learning platform.