

Automatic Parcel Damage Recognition Module for an Inspection Robot

Wiktor Goszczyński, Szymon Wałęga, Janusz Chmiel, Bartłomiej Gawęda, Katarzyna Grobler-Dębska

AGH University of Krakow

Dep.of Automatic Control and Robotics

al. Adama Mickiewicza 30

30-059 Kraków, Poland

Email: {wiktorg, swalega, januszchm}@student.agh.edu.pl, {bargaw, grobler}@agh.edu.pl

Abstract—This paper presents an AI-based solution for automated parcel damage detection, combining machine learning algorithms with a custom-built mobile inspection robot. A mobile robot was designed and built specifically for this task, equipped with a vision system and sensing components. We developed a dataset of over 6,800 images and applied a tailored data augmentation process to better capture the variability found in operational environments. Our approach refines a YOLOv11ncls-based model, achieving 98.50% accuracy, 97.% precision, and 99.74% recall on validation data. By optimizing the model for deployment on widely available hardware via CoreML, we reached inference speeds exceeding 251 FPS, ensuring rapid processing. An interactive dashboard was also created to monitor performance and facilitate comparisons between model iterations.

I. INTRODUCTION

MODERN warehouses are under increasing pressure to optimize logistics processes, which has led to a growing interest in automation technologies. Warehouse automation is evolving toward flexible and intelligent systems that leverage autonomous mobile robots (AMRs), artificial intelligence (e.g., for image analysis to detect damaged packages or support inventory, demand forecasting, robot path optimization, and automated quality control), and digital twins to optimize operations in real time. Increasing emphasis is being placed on human-machine integration, solution scalability, and the ability of systems to adapt to dynamic and unpredictable warehouse environments [13].

One promising application of these technologies is automated package inspection based on computer vision, which offers a scalable, consistent, and efficient alternative to manual quality control. Traditional methods for detecting damaged packages rely heavily on human inspection, which is time-consuming, inconsistent, and impractical in high-throughput environments. Automated inspection robots not only address these limitations but also serve as valuable platforms for testing and validating modern AI algorithms, effectively bridging the gap between theoretical research and practical deployment.

This synergy forms the foundation of the project initiated within the Industrial Data Science (IDS) student research

group. The main goal of this paper is to design and build a mobile inspection robot equipped with a computer vision system capable of detecting damaged packages in real time. This paper presents the process of constructing such a robot and implementing a computer vision algorithm for package damage detection. The contribution of this work is a functional prototype that demonstrates an AI-based package damage detection system intended for integration with a mobile robotic platform. The developed system achieves high classification accuracy and operates in real time.

The paper is organized as follows. The section I presents the context of warehouse automation and outlines the main objective of the project, which is to develop a parcel damage detection system. The section II reviews current solutions in computer vision, mobile robotics, and defect detection. The section III describes the construction and architecture of the developed mobile inspection platform. The section IV explains the process of data collection, exploration, and augmentation. In section V, the use of the YOLO model ("You Only Look Once"), training configuration, and iterative model improvement are discussed. The section V-B covers the conversion and preparation of the model for real-time inference. The results and evaluation section presents the model's performance metrics, error analysis, and results under challenging test conditions. The section VII outlines an interactive interface for evaluating and comparing model performance. The paper (sec. VIII, sec.IX) concludes with the future work and conclusion, which highlight potential directions for further development and summarize the achieved outcomes.

II. RELATED WORKS

Modern warehouses are increasingly adopting autonomous mobile robots (AMRs) to automate logistics processes. Unlike traditional automated guided vehicles (AGVs), AMRs can navigate dynamic environments and make independent decisions without fixed infrastructure [7]. Advances in sensing, control, and AI have enabled their use not only for material handling but also for inspection tasks [16]. A related study describes a cyber-physical warehouse system for order-picking using mobile robots, combining LiDAR for obstacle avoidance and optimized path-planning in dynamic environments [2]

This work was supported by AGH University of Krakow subvention for scientific activity no. 16.16.120.773 and by program "Excellence initiative—research university" for the AGH University no. 500.696.9400.

Several implementations showcase inspection capabilities. For instance, a MicroAir Vehicle with onboard SLAM and camera was used to scan QR codes in GPS-denied warehouses [9]. More advanced approaches integrate YOLO-based object detection for real-time condition assessment of packages and infrastructure [12]. Other systems combine LiDAR and depth cameras to support autonomous inventory validation [11].

Artificial intelligence plays a key role in these systems. A comprehensive review highlights the impact of AI and computer vision in enabling object recognition, quality control, and adaptive inspection [13]. Techniques like edge computing and reinforcement learning further enhance system responsiveness.

The YOLO (You Only Look Once) framework remains a leading architecture for real-time defect detection. Recent studies trace its development from v1 to v11, noting improvements like attention modules and deployment readiness on edge devices [5], [8]. In industrial settings, YOLO has been successfully applied to detect packaging defects [14], classify damage types with high accuracy [3], and improve connector inspection robustness [15].

Beyond warehouses, AI-powered visual inspection is explored in broader logistics, such as parcel sorting in rail transport systems [4].

The literature emphasizes the growing importance of advanced deep learning models, particularly YOLO, for real-time, accurate, and scalable inspection tasks in warehouse environments and autonomous mobile robots in tasks related to package control and inspection in warehouse environments. The integration of advanced sensors, vision systems, and AI algorithms with mobile robotic platforms lays a solid foundation for further developments toward fully automated warehouse processes and high-quality logistics services.

III. CUSTOM MOBILE INSPECTION ROBOT

A custom mobile inspection robot was developed as part of a student research initiative to support parcel damage recognition. The platform integrates key hardware components and provides mobility, sensing, and computation capabilities needed for logistics environments.

A. Design Overview

The robot was built with the following core requirements: compact, modular chassis for easy transport; omnidirectional mecanum wheels for maneuverability; and onboard sensors for robust environmental perception. The sensing suite includes an RGB camera, LiDAR, and ultrasonic distance sensors. Battery power ensures portability and extended runtime.

B. System Architecture and Integration

The robot is powered by two primary units: an NVIDIA Jetson Nano and an Arduino microcontroller. The Jetson handles camera input, sensor processing, and ML inference, while the Arduino controls motor operations. The RGB camera is mounted on a rotating servo for dynamic inspection, and LiDAR provides precise obstacle detection. Both units communicate via USB.

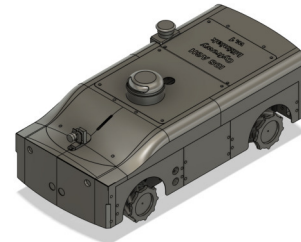


Fig. 1. Conceptual Model of the Inspection Robot

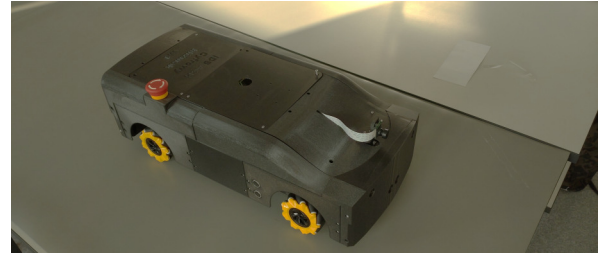


Fig. 2. 3D Printed Prototype of the Inspection Robot

C. Testing and Validation

The mechanical components were prototyped via 3D printing and tested in laboratory conditions. Functional validation confirmed smooth omnidirectional movement, stable sensor integration, and real-time data capture necessary for classification.

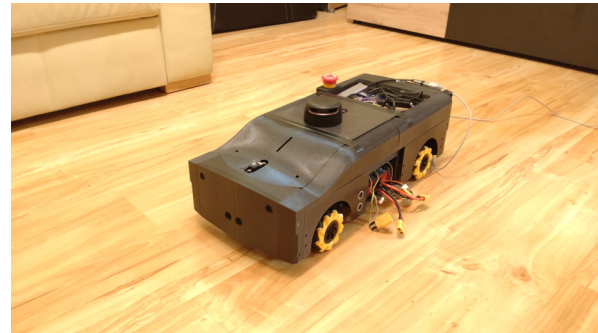


Fig. 3. Robot during LiDAR Testing in Lab Conditions

IV. DATASET DEVELOPMENT

The foundation of our parcel damage classification system developed for the inspection robot is a comprehensive, high-quality dataset that accurately represents the problem. This section details our approach to dataset construction, exploration, and augmentation.

A. Data Collection and Organization

Initial data collection began with a publicly available dataset from Kaggle [1] which provided a useful starting point but exhibited several limitations. We used the FiftyOne [10] framework for extensive dataset exploration and analysis. FiftyOne provides valuable tools for:

- Visualizing dataset samples and their distributions
- Identifying potential biases in the data
- Detecting and removing duplicate or near-duplicate images
- Verifying label quality and consistency

Upon inspection using this framework, we found several limitations of gathered photos:

- Poor image quality in numerous samples.
- Limited variety in package types and damage patterns.
- Presence of unrealistic or staged damage scenarios

To address these limitations, we supplemented the initial dataset with custom images captured in realistic environments in different situations and scenes. This approach improved the representativeness of the training data with respect to operational conditions encountered in logistics environments.

This exploration phase was crucial for ensuring dataset quality prior to model training and helped identify specific areas that required additional data collection.

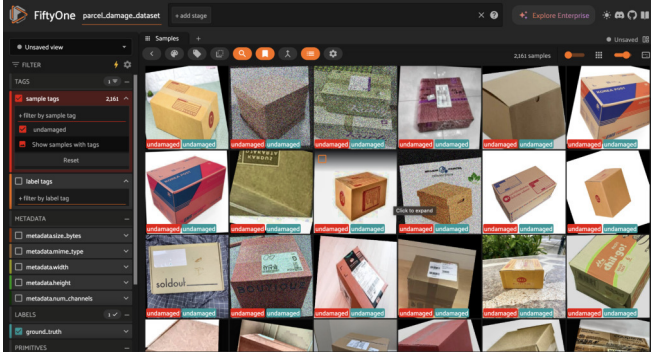


Fig. 4. Shows representative samples from our dataset, highlighting the variety of parcel types, lighting conditions, labeling, and damage patterns.

B. Augmentation Strategy

Data enhancement plays a critical role in improving the robustness of the model. Rather than applying generic augmentation techniques, we developed a domain-specific augmentation pipeline using the Albumentations library. Our pipeline was carefully designed to simulate realistic variations that parcels might encounter in logistics environments while avoiding unrealistic transformations that could harm model performance. We applied:

- Conservative spatial transformations to simulate realistic camera angle variations:
 - Slight rotations ($\pm 10^\circ$) with a 50% probability
 - Small affine transforms (scale factor 0.95-1.05, translation $\pm 3\%$)
 - Horizontal flips with a 30% probability
 - No-op spatial transformations with a 20% probability
- To simulate different indoor lighting conditions commonly found in warehouses and sorting facilities:
 - Random brightness/contrast adjustments ($\pm 10\%$) with a 50% probability

- Color temperature shifts (brightness $\pm 5\%$, contrast $\pm 5\%$, saturation $\pm 5\%$, hue $\pm 5\%$) with a 30% probability
- No lighting modifications with a 20% probability
- Shadow Effects To simulate natural lighting conditions:
 - Very mild random shadows with a 20% probability
 - No shadow effect with a 80% probability

The final pipeline included consistent resizing to 640×640 pixels while maintaining aspect ratio, implemented through a combination of LongestMaxSize and PadIfNeeded transformations with constant border padding. This pipeline was applied 10 times to each of the original photos.

The final dataset was organized into a structured format with two primary categories:

- Damaged: 2,372 training images and 903 validation images
- Undamaged: 2,600 training images and 1,000 validation images



Fig. 5. Illustrates examples of our augmentation pipeline applied to sample images, demonstrating the realistic nature of the transformations.

V. TRAINING CLASSIFIER

For our parcel damage detection task, we selected the YOLOv11n-cls model [6], a classification variant of the YOLO (You Only Look Once) family. This approach was chosen for several reasons:

- Efficient inference speed, critical for real-time applications in logistics
- Strong feature extraction capabilities from the backbone network
- Successful track record in related computer vision tasks
- Flexibility to be deployed on various hardware platforms

A. Training Methodology

The model was trained using the Ultralytics framework [6], which provides an efficient implementation of YOLO models with several optimizations for training and inference. The training configuration used the following parameters:

- Input image size: 640×640 pixels
- Training epochs: 100
- Model architecture: YOLOv11n-cls.pt

Other hyperparameters such as batch size, optimizer settings, and learning rate schedule were automatically determined by the Ultralytics framework based on dataset characteristics and hardware capabilities. This approach leverages the expertise embedded in the framework to achieve optimal training results without manual hyperparameter tuning reducing manual workload and streamlining the training process.

Our model development followed an iterative improvement process with following steps:

- 1) Initial training with the training part of created dataset resulted in our baseline model (best0)
- 2) Performance analysis of best0. That revealed a specific weakness: opened boxes were frequently misclassified as undamaged
- 3) Additional data collection focused specifically on opened boxes, labeled as damaged
- 4) Dataset augmentation with the new samples
- 5) Re-training with the enhanced dataset resulted in our improved model (best1)
- 6) Performance analysis of best1. That revealed no major weaknesses in classification

This iterative approach allowed us to improve the model performance. The figure 6 presents a visualization of the training

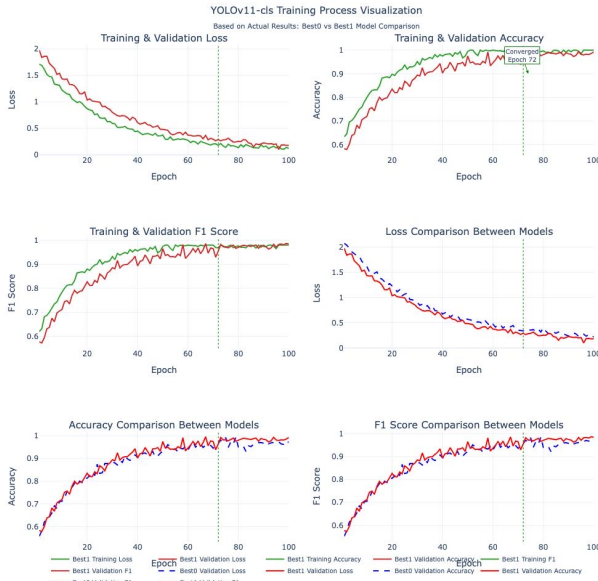


Fig. 6. Comparison of the training process between two YOLOv11-cl models (Best0 vs Best1) based on loss, accuracy, and F1 score for both training and validation datasets. Model Best1 demonstrates better convergence and higher classification performance overall.

process of the YOLOv11-cl classification model, comparing two model variants: Best0 and Best1. Based on the charts, Best1 consistently achieves better results than Best0 across all key metrics—loss, accuracy, and F1 score. Both accuracy and F1 score increase with each epoch, approaching a value of 1.0, while the loss steadily decreases. The convergence point occurs at epoch 72, after which the model performance

stabilizes. Throughout most of the training process, Best1 shows lower loss and higher accuracy and F1 score, indicating its superior classification performance compared to Best0.

B. Model Optimization

During testing with real-time inference the PyTorch model's frames per second (FPS) was around 30, meaning that it can process that number of sample pictures in one second. After converting to CoreML format the model reached 251+ FPS. We converted our models to ONNX format to address two primary objectives: first, to enable robust metric extraction and validation within our performance monitoring dashboard, as the standardized ONNX format provides more predictable parameter access patterns compared to native PyTorch. Secondly, changing model's format allowed us to establish a framework-agnostic model representation that enhances deployment flexibility and ensures consistent inference behavior across different runtime environments. The training process was monitored using metrics that included training loss, validation loss, and top-1 accuracy, which were recorded every epoch.

We achieved this using the following process:

- 1) Export of the trained PyTorch model to ONNX format
- 2) Conversion from ONNX to CoreML using the coremltools library
- 3) Optimization of the CoreML model for the Apple Neural Engine

The high inference speed makes the model suitable for integration into high-throughput logistics environments where real-time processing is critical.

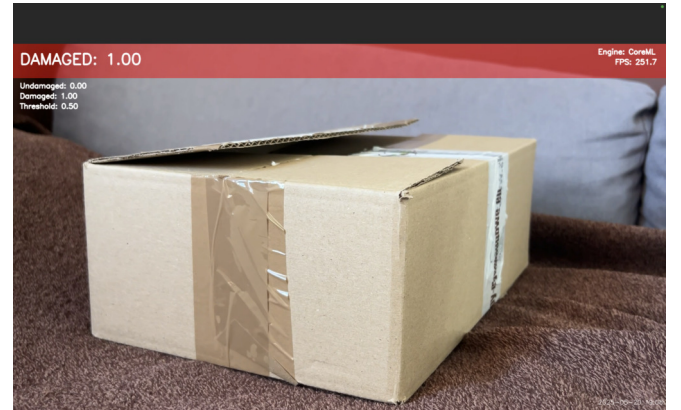


Fig. 7. Optimization process maintained model accuracy while significantly improving inference speed, achieving over 251 FPS on Apple M1 hardware during real-time testing.

VI. RESULTS AND EVALUATION

A. Performance Metrics

We evaluated our models using classification metrics including accuracy, precision, recall, and F1 score.

The best0 and best1 models achieved results as presented in Table I. The best1 model consistently outperformed the baseline (best0) across all evaluation metrics, indicating successful

mitigation of the identified dataset weaknesses. The enhancements in precision are particularly notable, indicating that our approach successfully reduced false-positive classifications.

TABLE I

SHOWS THE TRAINING STATISTICS FOR BOTH MODELS, ILLUSTRATING THE IMPROVEMENT IN FINAL PERFORMANCE ACHIEVED BY THE BEST1 MODEL AFTER ADDRESSING THE IDENTIFIED WEAKNESSES IN THE DATASET.

Metric	best0 model	best1 model	Improvement
Accuracy	97.00%	98.50%	+1.50pp
Precision	94.80%	97.04%	+2.24pp
Recall	98.81%	99.74%	+0.93pp
F1 Score	96.76%	98.37%	+1.61pp

We also evaluated our models under real-life conditions, such as too dark or too light environment. These results suggest that our models performance can be increased by hyper-tuning or changing the augmentation parameters for images. Scores achieved by both models in this scenario are presented in Table II.

TABLE II

SHOWS STATISTICS FOR BOTH MODELS IN BOTH DARKER AND LIGHTER ENVIRONMENTS USING REAL-LIFE IMAGES AND AUGMENTATIONS

Metric	best0 model	best1 model	Improvement
Accuracy	92.06%	94.44%	+2.38pp
Precision	89.74%	91.14%	+1.4pp
Recall	97.22%	100%	+2.78pp
F1 Score	93.33%	95.36%	+2.03pp

B. Confusion Matrix and Precision-Recall analysis

Detailed analysis of these matrices reveals:

- True Positives (Damaged correctly classified): Increased from 747 to 754
- False Negatives (Damaged classified as Undamaged): Decreased from 9 to 2, representing a 77.8% reduction
- True Negatives (Undamaged correctly classified): Increased from 871 to 889
- False Positives (Undamaged classified as Damaged): Decreased from 41 to 23, representing a 43.9% reduction

The substantial reduction in false negatives is particularly important for our application context. In parcel damage detection, false negatives (missed damages) typically have a greater business impact than false positives, as they can lead to damaged items being delivered to customers, resulting in customer dissatisfaction and potential financial losses.

C. Visual Inspection of Classification Results

Beyond quantitative metrics, we conducted a qualitative analysis of classification results to identify patterns in the remaining misclassifications. For the best1 model, we observed the following:

- False negatives (2 instances): Both cases involved parcels with minor damage that was partially obscured or at the edge of the package.

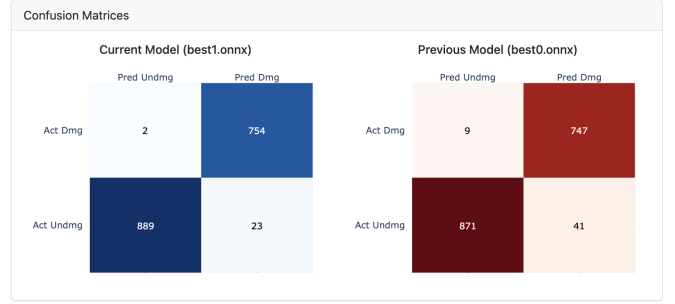


Fig. 8. Confusion matrices for both models

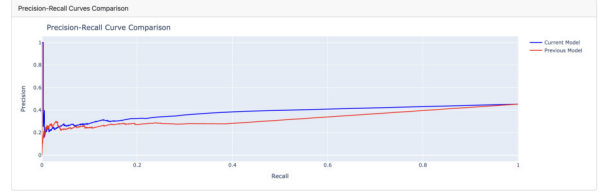


Fig. 9. Shows the precision-recall curves, where the best1 model (blue line) maintains higher precision across all recall values than best0 (red line). These analyses provide a more comprehensive view of model performance beyond the single operating point represented by the confusion matrix, demonstrating that the best1 model achieves superior performance across various classification thresholds.

- False positives (23 instances): Primarily involved parcels with heavy tape reinforcement or unusual packaging materials that resembled damage patterns.

These observations provide valuable insights for future improvements, suggesting that focusing on capturing more examples of minor/obscured damage and diverse packaging materials could further enhance model performance. Fig. 8

VII. DASHBOARD FOR VISUALIZATION AND ANALYSIS

To support model evaluation and facilitate comparisons, we developed a lightweight web-based dashboard using Dash and Plotly. The interface allows users to select between model versions (e.g., best0 vs best1), inspect performance metrics (accuracy, precision, recall, F1 score), and visualize confusion matrices and precision-recall curves. Users can also upload custom images or specify test folders for evaluation.

The dashboard supports real-time testing with side-by-side model predictions, along with confidence scores displayed via intuitive visual indicators. All visualizations are interactive, supporting threshold-independent assessments and flexible data exploration.

This interface enables rapid prototyping, debugging, and demonstration of classification results without requiring deep technical expertise. The tool has proven useful for both offline analysis and on-site testing scenarios.

VIII. FUTURE WORKS

Although the developed parcel damage classification system achieves high performance, several promising directions remain for further development, both in terms of visual analysis

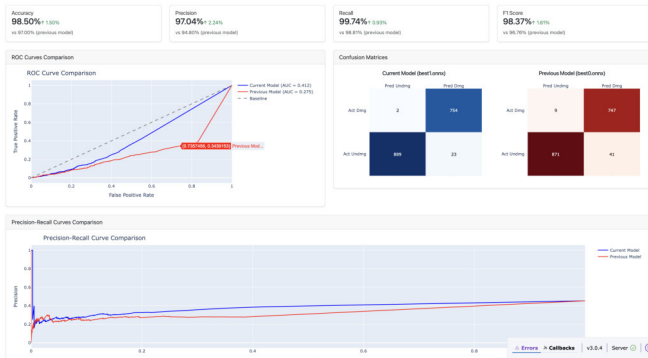


Fig. 10. Comprehensive dashboard presenting all the important information about currently selected model. It also give option for real-time image testing.

and integration with the mobile inspection platform. The key next steps include:

- 1) **Damage Localization:** Extending the system to not only classify parcels as damaged but also localize and highlight specific damaged areas.
- 2) **Damage Type Classification:** Developing a multi-class model that can distinguish between different types of damage (e.g., tears, water damage, crushing) to provide more detailed information.
- 3) **Integration with Mobile Robot:** Combining the visual inspection system with the custom-built mobile platform will enable fully autonomous parcel inspection within warehouse environments.
- 4) **Edge Deployment:** Further optimization for deployment on edge devices directly on conveyor systems using techniques like model quantization.
- 5) **Multi-View Integration:** Combining predictions from multiple camera angles to improve detection accuracy for damages that might only be visible from certain perspectives.
- 6) **Continuous Learning Framework:** Implementing a system for ongoing model improvement based on operational feedback and new data collection.

IX. CONCLUSION

This paper presents a complete parcel damage detection system based on machine learning, covering data preparation, dataset enrichment, model development, and real-time optimization. A key component of the solution is the custom-designed and built mobile inspection robot, tailored for use in warehouse environments.

The system achieved a very high recall rate (99.74%), demonstrating its ability to detect parcel damage reliably, an essential requirement in logistics to minimize customer complaints and reduce financial losses. The inference speed supports deployment in high-throughput environments, and the custom dashboard enables real-time performance monitoring and model comparison.

The integration of both components, the mobile robotic platform and the vision-based classification system, lays the

foundation for a fully autonomous, real-time parcel inspection solution. This project demonstrates that the thoughtful application of computer vision techniques, combined with domain-specific considerations and high-quality data, can yield practical and effective solutions to real-world challenges in logistics automation.

REFERENCES

- [1] Damaged and intact packages. <http://web.archive.org/web/20080207010024/http://www.808multimedia.com/winnt/kernel.htm>, accessed: 2025-04-23
- [2] Alhas, O.B., et al.: Towards developing a cyber-physical warehouse system for automated order-picking. In: *Annals of Computer Science and Information Systems*, Volume 32. pp. 321–328 (2022). <https://doi.org/10.15439/2022F208>
- [3] Chaudhary, K., Singh, A.: Parcel damage classification using computer vision: A deep learning approach for shipment quality assessment. In: *2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI)*. vol. 2, pp. 1–6 (2024). <https://doi.org/10.1109/IATMSI60426.2024.10503363>
- [4] Gupta, M., Garg, N., Garg, J., Gupta, V., Gautam, D.: Designing an intelligent parcel management system using iot & machine learning. In: *2022 IEEE IAS Global Conference on Emerging Technologies (GlobConET)*. pp. 751–756 (2022). <https://doi.org/10.1109/GlobConET53749.2022.9872449>
- [5] Hussain, M.: Yolo-v1 to yolo-v8, the rise of yolo and its complementary nature toward digital manufacturing and industrial defect detection. *Machines* **11**(7) (2023). <https://doi.org/10.3390/machines11070677>, <https://www.mdpi.com/2075-1702/11/7/677>
- [6] Jocher, G., Qiu, J.: Ultralytics yolo11 (2024), <https://github.com/ultralytics/ultralytics>
- [7] Keith, R., La, H.M.: Review of autonomous mobile robots for the warehouse environment (2024), <https://arxiv.org/abs/2406.08333>
- [8] Mao, M., Hong, M.: Yolo object detection for real-time fabric defect inspection in the textile industry: A review of yolov1 to yolov11. *Sensors (Basel)* **25**(7), 2270 (Apr 2025). <https://doi.org/10.3390/s25072270>
- [9] Martinez-Carranza, J., Rojas-Perez, L.O.: Warehouse inspection with an autonomous micro air vehicle. *Unmanned Systems* **10**(04), 329–342 (2022). <https://doi.org/10.1142/S2301385022410011>
- [10] Moore, B.E., Corso, J.J.: Fiftyone. GitHub. Note: <https://github.com/voxel51/fiftyone> (2020)
- [11] Parikh, H., Saijwal, I., Panchal, N., Sharma, A.: Autonomous mobile robot for inventory management in retail industry. In: Singh, P.K., Wierchoń, S.T., Chhabra, J.K., Tanwar, S. (eds.) *Futuristic Trends in Networks and Computing Technologies*. pp. 93–103. Springer Nature Singapore, Singapore (2022)
- [12] Sanchez-Cubillo, J., Del Ser, J., Martin, J.L.: Toward fully automated inspection of critical assets supported by autonomous mobile robots, vision sensors, and artificial intelligence. *Sensors* **24**(12) (2024). <https://doi.org/10.3390/s24123721>, <https://www.mdpi.com/1424-8220/24/12/3721>
- [13] Sodiya, E.O., Umoga, U.J., Amoo, O.O., Atadoga, A.: Ai-driven warehouse automation: A comprehensive review of systems. *GSC Advanced Research and Reviews* **18**(2), 272–282 (2024). <https://doi.org/10.30574/gscarr.2024.18.2.0063>
- [14] Vu, T.T.H., Pham, D.L., Chang, T.W.: A yolo-based real-time packaging defect detection system. *Procedia Computer Science* **217**, 886–894 (2023). <https://doi.org/10.1016/j.procs.2022.12.285>, 4th International Conference on Industry 4.0 and Smart Manufacturing
- [15] Wu, W., Li, Q.: Machine vision inspection of electrical connectors based on improved yolo v3. *IEEE Access* **8**, 166184–166196 (2020). <https://doi.org/10.1109/ACCESS.2020.3022405>
- [16] Xuan, O.W., Selamat, H., Muslim, M.T.: Autonomous mobile robot for transporting goods in warehouse and production. In: Tan, A., Zhu, F., Jiang, H., Mostafa, K., Yap, E.H., Chen, L., Olule, L.J.A., Myung, H. (eds.) *Advances in Intelligent Manufacturing and Robotics*. pp. 555–565. Springer Nature Singapore, Singapore (2024)