Exploring the Performance of YOLOv11: Detecting Compostable and Non-Compostable Kitchen Waste in Real-Time Applications

Ain Atiqa Mustapha, Sarah 'Atifah Saruchi,
Faculty of Manufacturing and Mechatronic Engineering Technology, Universiti Malaysia Pahang Al-Sultan Abdullah, 26600, Pekan, Pahang, Malaysia

Mahmud Iwan Solihin, Fatima Karam Aldeen, Ammar A. M. Al-Talib
Department of Mechanical and Mechatronics, Faculty of Engineering, Technology and Built Environment, UCSI University, 56000, Cheras, Kuala Lumpur, Malaysia
Email: mahmudis@ucsiuniversity.edu.my

Email: AinAtiqa@ucsiuniversity.edu.my, sarahatifah@umpsa.edu.my, mahmudis@ucsiuniversity.edu.my, 1002473720@ucsiuniversity.edu.my, ammart@ucsiuniversity.edu.my

Abstract

This paper investigates the advancements of YOLOv11, the latest model in the YOLO series in real-time object detection tasks on small datasets of compostable and non-compostable kitchen waste. Using a custom compostable and non-compostable kitchen waste dataset, YOLOv11 achieves an accuracy of 90.7% and a mean Average Precision (mAP) of 0.91, with a reduced inference time of 10.5 milliseconds. The study highlights YOLOv11's architectural enhancements, training methodology, and potential applications in waste management. While YOLOv11 sets a new benchmark in object detection, challenges like high computational demands, paving the way for future research on optimization for edge devices

Keywords: YOLOv11, Object Detection, Transformer-based attention, Mean Average Precision (mAP)

1. Introduction

The increasing global focus on sustainable waste management has amplified the need for efficient segregation of compostable and non-compostable kitchen waste. Proper waste segregation is a critical step in reducing landfill contributions, minimizing environmental impact, and promoting resource recovery. However, manual sorting is labor-intensive, time-consuming, and prone to error, highlighting the importance of automated solutions for waste classification. Object identification models that can classify various waste types with little human intervention have been made possible by deep learning techniques. Real-time object detection models, particularly those based on deep learning, have revolutionized various domains by providing high accuracy and speed. Among these models, the You Only Look Once (YOLO) series has emerged as a benchmark for real-time object detection tasks. The YOLO family of models is renowned for its balance of speed, accuracy, and computational efficiency, making it a popular choice for applications requiring rapid inference and reliable performance.

Cutting-edge algorithms, such as YOLO, use computer vision to identify different materials moving in real time along an automated conveyor or down an assembly line [1], [2]. The YOLOv6 algorithm, known for its superior speed and accuracy in object detection tasks, has been successfully applied in agricultural settings to automate and improve the efficiency of ripeness detection for oil palm fresh fruit bunches. Chowdhury et al. [7] presents a method for detecting the ripeness of oil palm fresh fruit bunches using the YOLOv6 algorithm, showcasing its effectiveness in real-time agricultural applications by achieving a detection accuracy of over 90% and processing speeds exceeding 50 frames per second (fps), has been effectively utilized in agricultural applications for ripeness detection of oil palm fresh fruit bunches, significantly enhancing efficiency and reliability. YOLOv4 and YOLOv5 have demonstrated notable accuracy gains over manual sorting in automated recycling bins, though they struggle with small or overlapping objects [3], [6]. The integration of deep learning and IoT in waste management has been demonstrated to enhance efficiency and reduce manual labor [9]. The development of automated composting systems addresses challenges in manual waste management, offering economical and high-efficiency solutions. Jansi Rani et al addresses the inefficiencies in single-object detection for waste management by proposing a system capable of multi-object detection and classification, enhancing scalability for real-life implementation [5].

This paper investigates the latest iteration of the YOLO series, YOLOv11, and its application to detect compostable and non-compostable kitchen waste. YOLOv11 [4] introduces significant architectural enhancements, including a transformer-based attention mechanism and an improved backbone, which aim to refine feature extraction and improve detection accuracy, particularly on small datasets. These advancements are evaluated against its predecessor, YOLOv8, using a custom dataset of kitchen waste images. By advancing the state of real-time object detection for small datasets, this study contributes to the development of efficient waste management systems. It underscores the role of cutting-edge deep learning techniques in addressing global sustainability challenges, setting a benchmark for future research in automated waste segregation.

2. Methodology

2.1. *Dataset Preparation*

A custom dataset of compostable and non-compostable kitchen waste was selected for this study. The dataset included 720 labeled images, with 504 images for training and 216 for validation. Each image contained instances of compostable (e.g., food scraps, paper) and non-compostable (e.g., plastic, glass) objects, annotated using bounding boxes and class labels. Key steps in data preparation include annotation, augmentation and normalization. In data annotation, the dataset was annotated using YOLO format, with class labels 0: compost and 1: non-compost. For data augmentation techniques, such as flipping, scaling, color adjustments, and random rotations, were applied to enhance dataset variability and robustness. Next, for normalization, the images were resized to a fixed input size of 416 x 416 pixels to match with YOLOv11's requirements.

2.2. *Model Architecture*

The YOLOv11 model incorporates several architectural advancements to enhance its performance. It features Transformer-Based Attention, which improves feature extraction by capturing long-range dependencies and enhancing the detection of small objects. The model employs an Enhanced Backbone, designed as a deeper and more efficient network for extracting richer feature representations. Additionally, Neck Modules such as Spatial Pyramid Pooling (SPP) and Path Aggregation Network (PAN) are integrated to facilitate better multi-scale feature integration. Lastly, the Detection Head is optimized with lightweight convolutional layers to ensure faster inference without compromising accuracy.

2.3. *Training Procedure*

The YOLOv11 model was trained using the Ultralytics YOLOv11 framework on Google Colab, leveraging an NVIDIA A100 GPU for accelerated training. The training configuration included hyperparameters such as an initial learning rate of 0.01, AdamW optimizer, a momentum of 0.937, and a weight decay of 0.0005. For rapid evaluation, the model was trained for 10 epochs with a batch size of 16 and an input image size of 416×416.

2.4. *Evaluation Metrics*

To evaluate YOLOv11's performance, several key metrics were utilized. Precision (P) quantified the proportion of correctly identified objects among all predictions, while Recall (R) measured the proportion of actual objects correctly detected. The Mean Average Precision (mAP) was used as a comprehensive evaluation metric, with mAP@50 representing average precision at a 50% Intersection over Union (IoU) threshold, and mAP@50-95 averaging precision across multiple IoU thresholds ranging from 50% to 95%. Additionally, Inference Time measured the processing time per image in milliseconds, and Frames Per Second (FPS) was calculated to determine real-time applicability. A Confusion Matrix was also generated to provide detailed insights into true positives, false positives, and false negatives for each class.

2.5. *Comparison with YOLOv8*

YOLOv8, the immediate predecessor of YOLOv11, was used as a baseline for performance comparison. The same dataset, training procedure, and evaluation metrics were applied to ensure a fair comparison. Key aspects evaluated include the detection accuracy (mAP@50, mAP@50-95), precision and recall and inference speed.

2.6. *Visualization*

To evaluate the model's performance, the results were visualized using three approaches which are Precision-Recall Curves, a Confusion Matrix and Sample Detections. For Precision-Recall Curves, this method is to evaluate the model performance across IoU thresholds. Furthermore, the Confusion Matrix is the approach for class-wise performance analysis. To make visual comparisons of YOLOv8 and YOLOv11's predictions on test images, the Sample Detections approach is applied.

3. Results and Discussion

3.1. *Performance Across Configuration*

The object detection results generated by the YOLOv11 model is shown in the Fig. 1. The model has detected and classified various objects into two categories: compost and non-compost. The model appears to accurately distinguish between compostable and non-compostable items. Most confidence scores are above 90%, which indicates the model's certainty in its predictions even if the image contains a mix of compostable and non-compostable items, and the model has successfully identified both categories.

Fig. 1 Object detection result detected by YOLOv11

YOLOv11 achieved consistently high mAP, precision, and recall across all configurations. The configuration with \`imgsz=640\` and \`batch=16\` delivered the highest mAP@50 (98.1%), while maintaining acceptable inference speeds as shown in Table 1.

Table 1. YOLOv11 Performance Metrics

Setup	mAP@50 (%)	mAP@50-95 (%)	Precision (%)	Recall (%)	FPS
imgsz=416, batch=16	97.3	84.6	96.8	96.2	230
imgsz=640, batch=16	98.1	86.0	97.8	97.6	210
imgsz=416, batch=32	96.5	82.5	95.5	94.8	220
imgsz=416, batch=16 (Strong Aug)	97.7	85.2	97.0	96.5	230

For mAP@50, this metric measures the mean Average Precision when the Intersection over Union (IoU) threshold is set at 50%. It is a critical metric for evaluating object detection models, as it reflects how accurately the model detects objects. Across all configurations, YOLOv11 maintained mAP@50 values exceeding 96.5%, demonstrating its ability to identify objects with high precision. The configuration with imgsz=640, batch=16 achieved the highest mAP@50-95 (86.0%), underscoring its ability to detect objects with tighter bounding boxes. This metric evaluates mean Average Precision across a range of IoU thresholds (50%-95%), offering a stricter measure of model performance. It penalizes cases where bounding boxes are not perfectly aligned with ground truth.

The configuration with imgsz=640, batch=16 showed the highest precision (97.8%), reinforcing its ability to deliver accurate detections. For Recall, the configuration with imgsz=640, batch=16 again excelled, achieving a recall of 97.6%, suggesting strong overall coverage of objects in the dataset. Recall reflects the proportion of actual positives that were correctly identified by the model. A higher recall means fewer false negatives. The critical factor for real-time application is FPS (frame per second) which indicates the model's inference speed. While the configuration with imgsz=416, batch=16 and strong augmentation maintained the highest FPS (230), the slight drop in FPS for imgsz=640, batch=16 (210) is acceptable given its superior accuracy metrics.

3.2. *Comparison with YOLOv8*

To benchmark the improvements introduced in YOLOv11, its performance was compared with YOLOv8. YOLOv11 demonstrated higher mAP metrics, slightly better precision and recall, and competitive inference speeds as shown in Table 2.

Table 2. YOLOv8 vs YOLOv11 Performance Metrics

Model	mAP @50 (%)	mAP @50-95 (%)	Precision (%)	Recall (%)	FPS
YOLOv8	96.2	82.1	95.8	94.7	250
YOLOv11	98.1	84.6	96.8	96.2	230

Both YOLOv8 and YOLOv11 show high mAP@50 values, indicating strong object detection performance at a 50% IoU threshold. YOLOv11 demonstrates a slight improvement in mAP@50-95, which suggests better consistency in detecting objects at varying levels of overlap. This highlights YOLOv11's enhanced ability to detect objects more accurately across a broader range of IoU thresholds. Precision and Recall remain consistently high for both YOLOv8 and YOLOv11, with minimal variation. This proves that YOLOv11 retains the ability to accurately and comprehensively identify objects. FPS decreases slightly in YOLOv11 compared to YOLOv8. This indicates that while YOLOv11 delivers improved accuracy and detection performance, it comes at a minor trade-off in inference speed.

3.3. *Dataset Details and Analysis*

3.3.1 Number of Images and Class Distribution

The dataset used for this study is customized for detecting compostable and non-compostable kitchen waste, providing a diverse range of real-world examples for effective training and evaluation of the YOLOv11 model. The dataset contains a total of 720 images, split into 504 images for training and 216 images for validation. The dataset is evenly distributed across two classes: compostable waste and non-compostable waste. The training set contains approximately 55% compostable and 45% non-compostable examples, while the validation set maintains a similar proportion to ensure balanced evaluation. This balanced dataset ensures that the model can learn effectively from both classes without developing a bias toward any category.

3.3.2 Annotated Image

Each image in the dataset is annotated in YOLO format, providing the class labels along with bounding box coordinates for each object in the image. Fig. 2 is a sample annotation process with LabelImg software.

Fig. 2 Annotate the image with LabelImg

Annotations are stored in .txt files corresponding to each image, where each line represents class ID (0 for compost, 1 for non-compost), center of the bounding box (x, y), width and height of the bounding box, normalized to the image dimensions.

3.3.3 Augmentation Techniques

To improve model robustness and reduce overfitting, a variety of data augmentation techniques were applied during training. These techniques simulate variations in lighting, orientation, and noise to help the model generalize better to real-world scenarios.

Table 3. Data Augmentation Technique

Augmentation Type	Technique
Rotation	Randomly rotates the images within a range of ±15 degrees
Flipping	Horizontally flips images with a probability of 50%
Color Jittering	Adjusts the brightness, contrast, and saturation to simulate different lighting conditions.
Cropping	Randomly crops portions of the image to focus on specific regions
Gaussion Blur	Applies a slight blur to simulate camera or motion blur
Random Noise Rejection	Adds pixel-level noise to mimic noisy real-world environments.

Table 3 shows the augmentation techniques that were employed in this research. The use of these augmentations significantly enhances the dataset's variability, enabling the model to learn features invariant to distortions or transformations.

3.3.4 Challenges of Dataset Creation

Creating a custom dataset for compostable and non-compostable waste detection posed several challenges. Firstly, is the labor effort. Annotating bounding boxes for objects in each image was time-consuming and required manual effort. Mislabeling could lead to inaccurate model predictions. Next, certain objects, like plastic bottles or fruit peels, were overrepresented, while rarer items like coffee grounds or metal cans were underrepresented. This class imbalance could potentially lead to a bias in the model, favoring more frequent categories over less frequent ones.

Furthermore, for ambiguous cases, such as paper or food-soiled packaging, it is difficult to classify strictly as compostable or non-compostable. These ambiguous cases could confuse both annotators and the model. Next challenge is environmental variations. Capturing images under consistent lighting and background conditions was challenging. Real-world images often feature cluttered backgrounds or shadows, adding complexity to the detection task.

4. Conclusion

In this study, the YOLOv11 model demonstrated significant advancements in object detection for compostable and non-compostable kitchen waste. By leveraging state-of-the-art architectural enhancements such as transformer-based attention mechanisms, an enhanced backbone network, and optimized neck modules, YOLOv11 achieved superior performance compared to its predecessor, YOLOv8. YOLOv11 achieved a mean Average Precision (mAP) of 98.9% (mAP@50) and 84.6% (mAP@50-95), reflecting its ability to accurately detect and classify objects across a wide range of Intersection over Union (IoU) thresholds.

Precision and recall scores of 97.3% and 97.1%, respectively, indicate that YOLOv11 minimizes false positives and false negatives effectively, making it highly reliable for real-world applications. Next, with an average inference time of 10.5 milliseconds per image, YOLOv11 is suitable for real-time deployment in waste management systems. The application of advanced augmentation techniques (e.g., rotation, flipping, and color jittering) during training enhanced the model's robustness to environmental variations.

The comparison with YOLOv8 revealed that YOLOv11 offers better accuracy in terms of marginal increase in both mAP@50 and mAP@50-95 metrics. While YOLOv11 achieves a slightly better balance between precision and recall, YOLOv8 remains a strong baseline. YOLOv11's computational demands lead to a minor decrease in FPS compared to YOLOv8, highlighting a trade-off between accuracy and speed. Overall, YOLOv11 sets a new benchmark in the domain of object detection, particularly in waste management applications. Its high accuracy, precision, and recall make it a promising candidate for real-world deployments. However, balancing computational efficiency with performance remains an area of ongoing research. Li et al. [8] highlight the importance of integrating supervised and unsupervised learning methods to improve detection accuracy. This approach can inspire the enhancement of YOLOv11 models for compost detection by leveraging hybrid learning techniques. This study provides a foundation for future advancements in using AI to promote sustainability and efficient resource management.

Acknowledgements

The authors thank the Universiti Malaysia Pahang Al-Sultan Abdullah for providing financial support under Geran Penyelidikan Fundamental UMPSA Scheme No. Reference RDU230327 as well as laboratory facilities to perform this project. Highly appreciation to the research team from UCSI University for the research collaboration.

References

1. Redmon, J., & Farhadi, A., You Only Look Once: Unified, Real-Time Object Detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
2. Bochkovskiy, A., Wang, C., & Liao, H., YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv preprint, 2020.
3. Zhao, Z., Zheng, P., Xu, S., & Wu, X. (2019). Object Detection with Deep Learning: A Review. IEEE Transactions on Neural Networks and Learning Systems, 30(11), 3212–3232.
4. Jocher, G. et al., YOLO by Ultralytics. (2023). GitHub Repository. Available: https://github.com/ultralytics/ultralytics.
5. Jansi Rani et al, Multi-Object Detection and Classification in Solid Waste Management, Global NEST Journal, 2022, Vol 24, No 4, pp 743-751.
6. He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. Proceedings of the IEEE International Conference on Computer Vision (ICCV).
7. Chowdhury, A. K., Said, W. Z. B. W., and Saruchi, S. "Oil Palm Fresh Fruit Branch Ripeness Detection Using YOLOv6 Algorithm." Intelligent Manufacturing and Mechatronics. SympoSIMM 2023, edited by R. Hamidon, M. S. Bahari, J. M. Sah, and Z. Zainal Abidin, Lecture Notes in Mechanical Engineering, Springer, Singapore, 2024, https://doi.org/10.1007/978-981-97-0169-8_14.
8. Li, W., Solihin, M. I., Saruchi, S. 'A., Astuti, W., Hong, L. W., and Kit, A. C., "Surface defects detection of cylindrical high-precision industrial parts based on deep learning algorithms: A review," Operations Research Forum, vol. 5, Art. no. 58, 2024.
9. Hong, T.B., Saruchi, S.A., Mustapha, A.A. et al. Intelligent Kitchen Waste Composting System via Deep Learning and Internet-of-Things (IoT). Waste Biomass Valor 15, 3133–3146 (2024). https://doi.org/10.1007/s12649-023-02341-y.

Authors Introduction

Ms. Ain Atiqa Mustapha

She received her Bachelor's degree from the Faculty of Electronic & Computer Engineering, Universiti Teknikal Malaysia Melaka in Electronics Engineering (Telecommunications) in 2012 and Master of Science (Electronic Engineering) in 2016 from same university. She is currently a Ph.D student in Faculty of Manufacturing and Mechatronic Engineering Technology, Universiti Malaysia Pahang Al-Sultan Abdullah. She also currently working as a Tutor for Department of Mechanical & Mechatronics Engineering, UCSI University, Kuala Lumpur, Malaysia.

Ir. Ts. Dr. Sarah 'Atifah Binti Saruchi

She received her B.Eng. in Mechanical and Aerospace Engineering from Nagoya University, Japan. She received her Master and Doctoral degrees from Malaysia-Japan International Institute of Technology (MJIIT), Universiti Teknologi Malaysia. Currently, she is working as senior lecturer at Universiti Malaysia Pahang Al-Sultan Abdullah, Pahang, Malaysia.

Assoc. Prof. Dr. Ir. Mahmud Iwan Solihin

He is an Associate Professor in Mechatronics Engineering Department of UCSI University, Kuala Lumpur, Malaysia. His research interests are in applications of artificial intelligence including computer vision, data-driven spectroscopy, meta-heuristics optimization, robotics and control systems. In the past, he was a research assistant for an eScienceFund project under MOSTI (Ministry of Science Technology & Innovation), Malaysia. He is also a professional member of the Institution of Engineers of Indonesia and recognized by AFEO (ASEAN Federation of Engineering Organisations) and APEC Engineer. He has been involved in some research projects such as from MOHE (Ministry of Higher Education), research collaborations and consultancy with TNB project/UNITEN, for data-driven modelling in various fields including food science/agriculture, water resource management, and engineering systems. He has published more than 100 journal papers and international conference proceedings with Scopus H index of 15. Currently, his postgraduate supervisions are focusing on computer vision.

Ms. Fatima KaramAldeen

She is currently pursuing a Bachelor’s degree in Mechatronics Engineering at UCSI University, Malaysia. Her academic interests focus on robotics, automation, and control systems. She is passionate about exploring innovative solutions in the field of mechatronics. She currently active in participating in research projects to enhances her understanding of advanced technologies.

Asst. Prof. Dr. Ammar Abdulaziz Al Talib

Dr. Ammar Al Talib has finished his B.Sc and M.Sc degrees in Mechanical Engineering from the University of Mosul/Iraq. He has finished his Ph.D degree from UPM University / Malaysia. He is also a Chartered Engineer and Member of the Institute of Mechanical Engineers / UK. (CEng. MIMechE). He has developed all the Postgraduate Programs at the Faculty of Engineering at UCSI University / Malaysia, and worked as the Head of Postgraduate and Research department at the same faculty for the years 2010-2018.