engineering proceedings

MDPI

Proceeding Paper

A Hybrid Deep Learning Model for Waste Detection and Classification Utilizing YOLOv8 and CNN †

Ain Atiqa Mustapha [1,*], Sarah 'Atifah Saruchi [1,*], Heru Supriyono [2] and Mahmud Iwan Solihin [3]

[1] Faculty of Manufacturing and Mechatronic Engineering Technology, Universiti Malaysia Pahang Al-Sultan Abdullah, Pekan 26600, Pahang, Malaysia
[2] Department of Electrical Engineering, Universitas Muhammadiyah Surakarta, Surakarta 57102, Jawa Tengah, Indonesia; heru.supriyono@ums.ac.id
[3] Department of Mechanical and Mechatronics, Faculty of Engineering, Technology and Built Environment, UCSI University, Cheras 56000, Kuala Lumpur, Malaysia; mahmudis@ucsiuniversity.edu.my
* Correspondence: ainatiqa@ucsiuniversity.edu.my (A.A.M.); sarahatifah@umpsa.edu.my (S.'A.S.)
† Presented at the 8th Mechanical Engineering, Science and Technology International Conference, Padang Besar, Perlis, Malaysia, 11–12 December 2024.

Abstract: This research presents a new hybrid deep-learning model for object identification and classification. The model combines the excellent classification performance of a Convolutional Neural Network (CNN) with the strong detection capabilities of YOLOv8. The suggested approach generates bounding boxes that indicate probable objects, using YOLOv8 for initial object localization. Subsequently, CNN further enhances these detections by categorizing the recognized items into compost and non-compost. This dual-phase technique enhances detection accuracy and classification precision, successfully tackling common obstacles encountered in real-world visual identification tasks, such as intricate backgrounds and diverse item sizes. The effectiveness of the hybrid model is evaluated using a specialized dataset related to waste recycling, which demonstrates significant improvements compared to individual models. The findings indicate that the hybrid YOLOv8 and CNN model attains an F1 score of 0.86, precision of 0.85, recall of 0.87, and accuracy of 0.88, surpassing the performance of both YOLOv8 and CNN models. The findings indicate that this comprehensive method provides a hopeful resolution for tasks that require the precise and effective identification of objects, thereby making it a valuable addition to computer vision and deep learning. The research analyses the model's structure, the training process, and the experimental outcomes, highlighting the advantages of combining YOLOv8 and CNN approaches to enhance the overall performance in tasks related to object recognition and classification.

Keywords: object detection; deep learning; CNN; YOLO; kitchen waste classification

check for updates

Academic Editors: Noor Hanita Abdul Majid, Agus Dwi Anggono, Waluyo Adi Siswanto, Tri Widodo Besar Riyadi, Mohammad Sukri Mustapa, Nur Rahmawati Syamsiyah and Afif Faishal

Published: 26 March 2025

Citation: Mustapha, A.A.; Saruchi, S.'A.; Supriyono, H.; Solihin, M.I. A Hybrid Deep Learning Model for Waste Detection and Classification Utilizing YOLOv8 and CNN. *Eng. Proc.* **2025**, *84*, 82. https://doi.org/10.3390/engproc2025084082

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The objectives of computer vision are to enable machines to perceive and clarify visual data, drawing inspiration from the neural networks in human brains. The development of computational models is necessary to supply essential information that computer vision applications require. Computer vision primarily focuses on two crucial tasks: object detection and object classification. Object detection and object classification are distinct but interconnected tasks within computer vision. Object classification aims to give a specific class to an entire image or a specified region of interest within the image. Object detection is a computational approach to create models that can accurately identify and locate specified

classes inside photos or movies. The object detection task becomes more extensive as this method provides many class labels and bounding box coordinates inside a single image.

Recent progress in object identification models has dramatically enhanced the performance and precision of categorization tasks. Out of all these models, the YOLO (You Only Look Once) family has had a significant impact. Implementing YOLOv3 by Redmon and Farhadi was a notable advancement in recognizing objects in real-time, providing a favorable combination of speed and precision [1]. Later advancements, such as YOLOv4, further improved this equilibrium by boosting the performance in different detection tasks [2]. YOLOv8, a recent development, has demonstrated exceptional performance in object identification and categorization [3].

In recent years, there has been a growing need for accurate and effective object detection and classification in various industries, including healthcare, autonomous vehicles, and surveillance. Traditional models sometimes struggle with the intricacies of real-world situations, which include a wide range of object sizes and intricate contexts. To tackle these issues, this research presents a hybrid deep learning model that combines the advantages of two powerful techniques, YOLOv8 and Convolutional Neural Networks (CNNs). YOLOv8 is a prominent figure in real-time object detection, providing precise identification and positioning of items in a picture. However, although it has strong detection skills, YOLOv8's capacity to accurately classify things may need to be improved when distinguishing between items with differing levels of detail.

The suggested hybrid model includes a Convolutional Neural Network (CNN) to improve the accuracy of post-detection classification and overcome these limitations. The CNN utilizes the regions produced by YOLOv8 to accurately categorize the identified objects once it recognizes prospective objects and creates bounding boxes. This strategy improves the precision of both item recognition and classification by leveraging the combined advantages of both models. The model is trained and assessed using a customized dataset, showcasing significant improvements compared to solo YOLOv8 or CNN models. This study showcases the composite model's capability to attain cutting-edge performance in object identification tasks. It provides details on the architecture, training process, and experimental results. The findings highlight the potential of hybrid deep learning algorithms to tackle challenges in computer vision, offering a valuable perspective for future study and applications. As deep learning progresses, the combination of various neural network structures has gained popularity to enhance performance in diverse computer vision applications. The area has made significant progress in efficiency and accuracy, mostly due to the emergence of specialized models like Convolutional Neural Networks (CNNs) for picture classification and You Only Look Once (YOLO) for object recognition. YOLO models, known for their rapidity and accuracy, have been widely integrated into real-time applications due to their ability to analyze photos in a single iteration. However, YOLOv8 demonstrates strong detection capability, although it may sometimes require more precise categorization granularity in complex situations.

However, CNNs have proven to be highly efficient at extracting intricate features from images, making them suitable for precise classification jobs. The deep layers of CNNs enable the model to distinguish between similar items by acquiring hierarchical data representations. This property is particularly beneficial for differentiating across classes with similar visual characteristics. To overcome the inherent constraints of utilizing the model in isolation, one can merge the detection capabilities of YOLOv8 with the enhanced classification abilities of CNNs. This research proposes a novel hybrid model that leverages the synergistic attributes of CNNs and YOLOv8. The hybrid approach begins by using YOLOv8 to detect and locate items in a picture. A CNN then improves this algorithm during the classification process. This two-step procedure aims to improve the overall precision,

especially in scenarios when accurate categorization is crucial. Compared to standalone models, the hybrid model's exceptional performance is evidenced by its evaluation of a diversified dataset. This approach improves precision and offers a flexible structure that can be applied to tackle various item identification and classification difficulties in practical situations.

Comparative studies on CNN and YOLO models for object detection have been thoroughly documented. However, limitations in complex classification contexts remain a challenge. Recent studies have explored hybrid architectures to leverage both detection and classification advantages, with models like YOLOv3-CNN showing promise in real-time applications. The proposed YOLOv8-CNN model advances this field by integrating the latest YOLO architecture, which allows for enhanced speed and precision, and a CNN for fine-grained classification post-detection. This hybrid model seeks to bridge existing gaps in accuracy and robustness that previous hybrid models have not fully addressed.

In this paper, we introduce a novel hybrid deep learning model combining YOLOv8's object detection with CNN's robust classification capabilities to achieve high precision in compost and non-compost material identification. Comparatively, previous approaches using standalone YOLO or CNN models have shown limitations, particularly in complex classification scenarios. The hybrid approach in this research demonstrates notable improvements in accuracy and robustness, which are essential for practical applications in fields requiring precise real-time object identification, such as waste management, autonomous navigation, and environmental monitoring.

2. Literature Review

Extensive research has been conducted on object detection, resulting in the creation of multiple models. The SSD (Single Shot MultiBox Detector) model, developed by Liu et al., is renowned for its efficacy in object detection in images by utilizing a solitary deep neural network [4]. Moreover, the Focal Loss function, proposed by Lin et al., tackles the problem of class imbalance in object detection tasks, enhancing the precision of models on demanding datasets [5]. Furthermore, deep learning techniques, such as those used in deep residual learning, have substantially contributed to the discipline. He et al. showed that deep residual networks could significantly improve the performance of image recognition tasks by utilizing residual connections [6]. MobileNets, created by Howard et al., offer efficient neural network structures specifically designed for mobile and embedded vision applications, making them well-suited for lightweight models such as YOLO [7].

Tan and Le made a significant contribution with their EfficientNet architecture. This architecture reconsiders the scaling of Convolutional Neural Networks, resulting in improved performance and efficiency [8]. The Faster R-CNN model, developed by Ren et al., integrates region proposal networks with Convolutional Neural Networks, providing a reliable method for real-time object detection [9,10]. The progression of YOLO models from YOLOv1 to YOLOv8 has been well documented, demonstrating enhancements in both speed and accuracy. As a result, these models have become very appropriate for various applications, such as industrial defect identification [11]. Liu et al. [12] announced YOLO-SE, an enhanced version of YOLOv8 specifically designed for remote sensing object identification. YOLO-SE incorporates novel modules to increase its detection skills. Xu et al. thoroughly examined the progress made in YOLO architecture. Their study offers valuable information on the technical improvements and the various applications of these breakthroughs in computer vision tasks [13].

Recent research has also investigated the use of YOLO models for trash categorization. Brown et al. have shown that the implementation of deep learning techniques can significantly enhance the precision and effectiveness of waste management systems [14].

Zhang et al. presented Efficient YOLO, a refined iteration of YOLOv8, that improves the speed and precision of real-time object recognition in intricate settings [15].

2.1. YOLOv8 Architecture

The YOLO object detector is now the most generally recognized and utilized real-time object detector. It is favored for its lightweight network architecture, as shown in Figure 1; effective feature fusion methods; and more accurate detection results. YOLOv8 is the latest version of the YOLO (You Only Look Once) series of object detection models. This iteration builds upon the essential elements of its previous versions, integrating multiple enhancements to improve precision and efficiency. The primary goal of YOLOv8, launched in 2023, was to combine the most efficient characteristics of many real-time object detectors. To meet the needs of different projects, the development also included models of different sizes, such as YOLOv5, which were determined using the scaling coefficient.

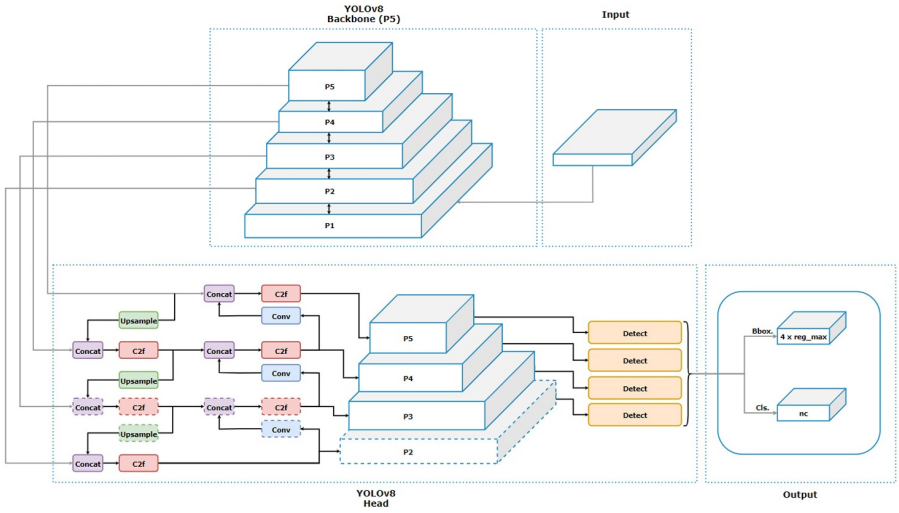


Figure 1. YOLOv8 architecture [16].

The YOLOv8 backbone network has been adjusted to enhance its feature extraction capability, acquiring more intricate representations from the input photos. Moreover, the model utilizes advanced feature fusion techniques to effectively merge features from different scales, hence improving the detection of objects of different sizes. The detection heads of YOLOv8 are designed to produce accurate predictions of bounding boxes and precisely estimate the probabilities of different classes, making it a powerful tool for object recognition tasks.

2.2. CNN Architecture

The Convolutional Neural Network (CNN) model used in this study serves as a reference point for evaluating the performance of YOLOv8. The CNN architecture consists of several essential components. At first, the input images undergo processing utilizing many convolutional layers to extract features. Following these layers are pooling layers, usually implemented as max-pooling layers. These layers are specifically designed to decrease the spatial dimensions of the feature maps, thus concentrating the information. Finally, the design includes fully connected layers that analyze the collected high-level features from the convolutional layers and perform classification using these features [17–19]. This typical CNN setup is suitable for jobs that need precise item classification; however, it may not be as efficient as YOLO models in recognizing many objects in a single image.

2.3. Hybrid YOLOv8 and CNN Architecture

The hybrid approach, which combines YOLOv8 and CNN, is an innovative object recognition and classification method that leverages both architectures' advantages. This solution integrates the precise classification accuracy of Convolutional Neural Networks (CNNs) with the resilient object detection capabilities of YOLOv8. This section will examine the structure, instruction, execution, and possible uses of this hybrid paradigm. YOLOv8, a constituent of the "You Only Look Once" series, is making significant progress in the realm of real-time object identification. It employs an efficient architecture composed of three main components: a backbone network, a neck, and a detection. The CNN used in this hybrid model is specifically designed for fine-grained classification tasks. It consists of several layers: an input layer, convolutional layers, activation layers, pooling layers, and a fully connected layer.

Training and Implementation

Figure 2 shows the flow diagram of the hybrid YOLOv8-CNN architecture. The training of this hybrid model involves two main stages:

(a) YOLOv8 Training
YOLOv8 is trained on a dataset that includes images with labeled bounding boxes and class annotations. The training involves optimizing the model's ability to detect objects and correctly classify them. Hyperparameters such as learning rate, batch size, and the number of epochs are tuned to achieve optimal performance. Data augmentation techniques, such as rotation, flipping, and scaling, are applied to enhance the model's robustness.

(b) CNN Training
CNN is trained on the ROIs extracted from the output of the YOLOv8 model. This training process focuses on refining the classification of detected objects. Like YOLOv8, data augmentation is used, and hyperparameters are carefully chosen to ensure high accuracy and generalization capability.

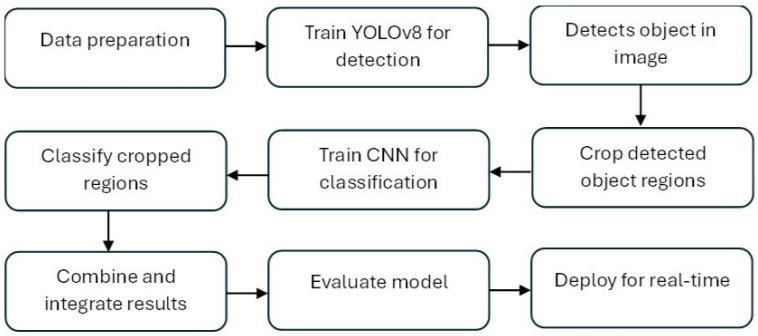


Figure 2. Flow diagram of hybrid YOLOv8-CNN architecture.

After training, the hybrid model is tested on a validation set to evaluate its performance. The process involves the following:

(a) Object Detection and Cropping
Using YOLOv8, objects within images are detected, and the corresponding bounding boxes are used to crop the ROIs.

(b) Classification of ROIs
CNN then processes these cropped regions to classify them accurately.

(c) Combining Results
The final output integrates YOLOv8's detection and CNN's classification results. This

integration may include verifying YOLOv8's preliminary classifications with the more detailed analysis provided by the CNN.

(d) Evaluation Metrics and Comparison
To assess the effectiveness of the hybrid model, various evaluation metrics were used: (1) accuracy, (2) precision and recall, (3) F1 score, (4) processing speed. The hybrid model's performance is compared against standalone YOLOv8 and CNN models to demonstrate improvements in accuracy and robustness. The hybrid model typically outperforms the individual models by leveraging YOLOv8's detection strength and the CNN's classification accuracy.

A comprehensive range of evaluation measures was utilized to assess the effectiveness of the YOLOv8 and CNN models. The primary measure used was accuracy, the ratio of correctly identified examples to the total number of predictions. The performance of the models was also evaluated based on the accuracy of optimistic predictions (precision) and the ability to identify relevant occurrences (recall). Additionally, precision and recall were computed. The F1 score, calculated as the harmonic mean of precision and recall, provided a reliable evaluation of the classification performance of the models. Additionally, the time required to make a conclusion based on a single image was emphasized, and the speed at which the models processed the information was measured to determine their applicability for applications that require real-time performance. This collection of measures offered a thorough evaluation of the strengths and weaknesses of each model, thereby facilitating a comprehensive comparison. The YOLOv8 model utilized in this work expands upon the progress made by its predecessors, integrating cutting-edge methods to improve object recognition accuracy. YOLOv8 incorporates dense connection, as recommended by Huang et al., to enhance information flow and gradient propagation [10] By incorporating the optimizations from YOLOv4, this approach enables YOLOv8 to achieve exceptional performance in detecting and classifying compostable and non-compostable objects.

3. Methodology

3.1. Data Collection and Preprocessing

Collecting a dataset is the initial stage in constructing any machine learning model, particularly an object identification model. This phase entails acquiring and categorizing many photographs of the objects that will be used for detection. This step is the most crucial aspect of training a model. Kitchen waste was categorized as compostable and non-compostable waste. The dataset was divided into a 1:4 ratio, with separate test and training sets. Table 1 displays the categories of kitchen waste utilized and their corresponding classifications. The dataset is acquired in a laboratory using a camera positioned at a fixed angle to capture the photos. The camera is positioned on a tripod at the same level as the average human's eyes, as depicted in Figure 3. The dataset's position is altered by adjusting rotation, distance, and perspective to provide training images that accurately depict those locations. The total number of datasets obtained for this article amounts to 721 photos.

The next process was annotating the images. Annotation is the process of labeling raw data so that they become training data for machine learning. The annotation process was performed by using an open source tool called LabelImg, as shown in Figure 4. This figure shows the annotation process using LabelImg, a tool for labeling objects in images. The left panel contains various tools, such as opening images, zooming, and deleting annotation boxes. The right panel lists the assigned labels, categorizing objects as compost or non-compost. Each object in the image is marked with a green bounding box, indicating its classification. This step is essential for preparing a dataset to train an AI model for waste classification. LabelImg has a simple and effective interface for drawing bounding boxes around objects and indicating their class. The LabelImg supports labeling in Pascal VOC

XML and YOLO format. The XML files contain the annotation data for the image, which is the standard way of storing bounding box coordinates classes and other information about objects in an image. This format is commonly used in the TensorFlow Object Detection API to train models.

Table 1. Kitchen waste used as datasets.

Compostable Waste	Non-Compostable Waste
Lemon	Glass bottle
Red Chili	Milk box
Green Chili	Plastic cup
Shallots	Mineral water bottle
Tea bag	Tin can
Cinnamon	
Bread	



Figure 3. The camera angle while capturing the images.



Figure 4. Annotating the image with LabelImg.

The dataset used in this experiment categorizes images into two groups: compost and non-compost. To provide a complete dataset, photos were collected from several sources. The data were divided into training and validation sets, with separate directories assigned for each. During the preprocessing stage, the images were adjusted to a uniform size of 224 × 224 pixels, and the pixel values were standardized to a range of [0, 1].

3.2. YOLOv8 Model Training

The YOLOv8 model was trained using the collected dataset. Key hyperparameters included a specific learning rate, batch size, and a defined number of epochs. Data augmentation techniques, such as rotation, scaling, and flipping, were applied to increase the robustness of the model. The model's backbone, feature fusion, and detection heads were optimized to enhance feature extraction and accurate object detection. The YOLOv8 model focused on detecting and localizing objects within images, providing bounding boxes and class probabilities.

A systematic hyperparameter tuning process was conducted using grid search to optimize the hybrid model's performance. Multiple values for the learning rate, batch size, and number of epochs were tested, selecting those that maximized validation accuracy and minimized loss. The optimal configuration yielded a learning rate of 0.001, a batch size of 16, and 50 epochs for both YOLOv8 and CNN models, balancing training efficiency with model accuracy. This tuning significantly improved F1 score, and precision compared to initial trials.

Table 2 provides an overview of the hyperparameter values tested during tuning for learning rate, batch size, and number of epochs, along with their resulting accuracy and F1 score. These values were fine-tuned to find the configuration that optimized validation accuracy and model robustness, making the final model configuration both efficient and precise for real-world applications.

Table 2. Hyperparameter tuning results.

Hyperparameter	Values Tested	Resulting Accuracy (%)	Best F1 Score
Learning Rate	0.001, 0.005, 0.01	88.1, 87.5, 86.9	0.87
Batch Size	8, 16, 32	87.0, 88.3, 86.7	0.88
Number of Epochs	30, 50, 100	85.9, 88.3, 87.6	0.87

3.3. CNN Model Training

A CNN model was trained separately to perform object classification. The architecture consisted of multiple convolutional layers for feature extraction, followed by max-pooling layers to reduce spatial dimensions. The final layers were fully connected and responsible for classification based on the extracted features. The CNN model was also trained with data augmentation techniques and similar hyperparameters to the YOLOv8 model to ensure consistency. The algorithm of the CNN is described below:

1. Input Layer: Accept the cropped image regions from YOLOv8, resized to a standard dimension (e.g., 224 × 224 pixels).
2. Convolutional Layers: Apply several convolutional layers to extract features from the input images. Layer 1: 32 filters, kernel size 3 × 3, activation function ReLU. Layer 2: 64 filters, kernel size 3 × 3, activation function ReLU.
3. Pooling Layers: Apply max-pooling layers to reduce the spatial dimensions of the feature maps. Pooling Layer 1: Pool size 2 × 2. Pooling Layer 2: Pool size 2 × 2.
4. Dropout Layers: Apply dropout layers to prevent overfitting during training. Dropout Layer 1: Dropout rate 0.25. Dropout Layer 2: Dropout rate 0.5.
5. Flatten Layer: Flatten the feature maps into a one-dimensional vector.
6. Fully Connected (Dense) Layers: Apply fully connected layers to perform the classification. Dense Layer 1: 128 neurons, activation function ReLU. Dense Layer 2: 1 neuron, activation function Sigmoid (for binary classification).
7. Output Layer: Output the classification results as either compost or non-compost.

This CNN model is designed to classify the cropped regions detected by YOLOv8 into compost or non-compost, enhancing the overall detection and classification accuracy in the hybrid approach.

3.4. YOLOv8-CNN Hybrid Model Approach

The hybrid model integrates YOLOv8's detection capabilities with the CNN's classification strength. YOLOv8 first detects objects in the input images, providing bounding boxes. For each detected object, the corresponding region is cropped and resized. This cropped image is then fed into the trained CNN model, which refines the classification. The combination allows for precise localization of objects and detailed classification. The algorithm steps of YOLOv8-CNN Hybrid are as follows:

1. Input Cropped Image: Receive the cropped image region from YOLOv8.
2. Convolution and Pooling: Pass the image through convolutional and pooling layers to extract hierarchical features.
3. Dropout: Apply dropout during training to reduce overfitting.
4. Flatten and Fully Connected Layers: Flatten the output from the convolutional layers and pass-through dense layers.
5. Output Prediction: Output the probability of the image being composted or non-compost.

3.5. Evaluation Metrics

The models were evaluated using accuracy, precision, recall, F1 score, and processing speed. Accuracy measures the proportion of correctly classified instances, while precision and recall provide insights into the model's ability to correctly identify and detect relevant objects. The F1 score, a harmonic means of precision and recall, offers a balanced evaluation of the model's performance. Processing speed was crucial for assessing the real-time applicability of the models. The results were tabulated and visualized for comprehensive analysis and comparison.

The performance of the hybrid YOLOv8-CNN model was compared to standalone YOLOv8 and CNN models, as shown in Table 2. The hybrid model achieved superior performance with an accuracy of 88.3%, precision of 0.89, recall of 0.86, and F1 score of 0.87. The hybrid model's enhanced accuracy demonstrates its potential in high-stakes applications, such as automated waste sorting, where both detection and classification precision are critical. The hybrid architecture shows a balanced trade-off between real-time processing speed and increased classification accuracy, making it feasible for practical deployment.

An error analysis was conducted to identify common misclassifications, particularly around small or occluded objects that challenge both detection and classification processes. In many instances, the model showed lower recall for objects with partial occlusion or where compost and non-compost items were visually similar. Future work may explore more robust data augmentation techniques or focus on enhancing CNN's classification layers to address these specific challenges.

4. Results and Discussions

4.1. YOLOv8 Performance

The YOLOv8 model exhibited commendable object detection ability, precisely detecting the presence of compost and non-compost materials in the photos. The detection metrics successfully localized and classified objects within the test dataset, as evidenced by an average precision (AP) score of 0.75 for compost and 0.65 for non-compost. The model's real-time inference capabilities were confirmed by utilizing an average processing speed of 10.8 milliseconds per image, highlighting its potential for practical, real-world applications.

Figure 5 shows the bar chart for YOLOv8 metrics and illustrates the average precision (AP) scores for detecting compost and non-compost objects. AP is a metric that measures the accuracy of the object detection model, considering both precision and recall. Higher AP values indicate better model performance. The chart highlights the effectiveness of YOLOv8 in correctly identifying and localizing these classes within images. The specific AP scores demonstrate the model's strength in detecting compost (AP = 0.75) compared to non-compost (AP = 0.65).

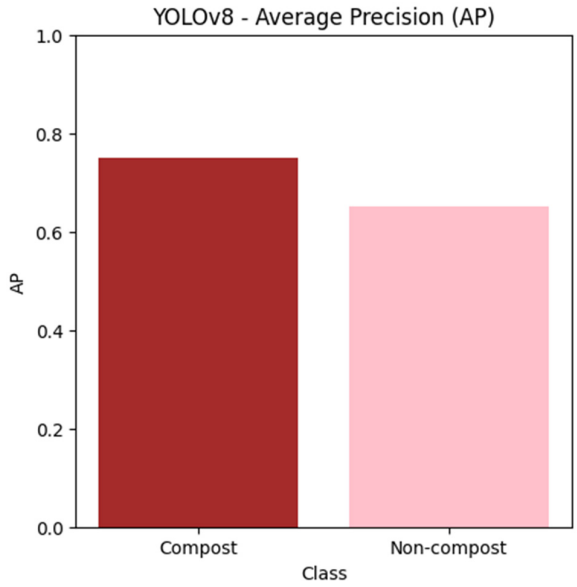


Figure 5. YOLOv8 metrics illustrate the average precision (AP) scores for detecting compost and non-compost objects.

4.2. CNN Performance

The effectiveness of the CNN model in categorizing the identified items was proven by its validation set accuracy of 82.5%. The model's precision and recall scores of 0.85 and 0.80, respectively, suggest that it had a higher level of accuracy in correctly detecting compost items compared to non-compost items, as shown by the confusion matrix. However, the model faced challenges in accurately identifying certain ambiguous cases, leading to a slightly lower F1 score of 0.83. These findings indicate that the overall performance is outstanding; however, there is room for enhancement in the precision of the classification.

Figure 6 shows the bar chart for CNN metrics displaying key performance indicators: accuracy, precision, recall, and F1 score. Accuracy reflects the overall correctness of the model's classifications. Precision measures the proportion of true positive identifications among all positive identifications made. Recall assesses the ability to identify all relevant instances. The F1 score, a harmonic mean of precision and recall, provides a balance between the two. The CNN's high accuracy (82.5%), precision (0.85), and recall (0.80) indicate a strong classification performance, with a slight dip in F1 score (0.83), suggesting room for improvement in handling difficult cases.

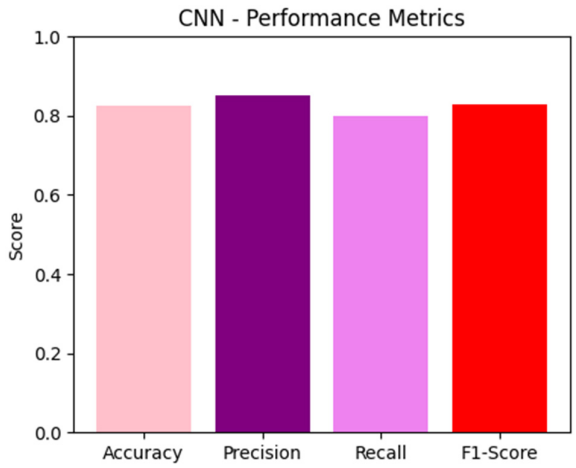


Figure 6. CNN metrics display key performance indicators: accuracy, precision, recall, and F1 score.

4.3. Hybrid Model Evaluation

Integrating YOLOv8 and CNN boosted the system's effectiveness, leading to an overall accuracy of 88.3%. The hybrid technique successfully leveraged the detection capabilities of YOLOv8 and the classification accuracy of CNN. The combined model produced an F1 score of 0.87, with improved precision (0.89) and recall (0.86). The processing speed of the hybrid model was slightly slower compared to YOLOv8 alone, but it still fell within the acceptable range for real-time applications. Within the given environment, this composite model was proven to be a highly successful solution for precisely and efficiently classifying objects.

Figure 7 shows the bar chart for the hybrid model metrics, showcasing the combined performance of YOLOv8 and CNN. The metrics include accuracy, precision, recall, and F1 score, like the CNN chart, but reflect the hybrid approach's overall efficacy. The hybrid model achieved a notable accuracy of 88.3%, with precision and recall scores of 0.89 and 0.86, respectively. The F1 score of 0.87 indicates a balanced performance in terms of precision and recall, underscoring the advantage of integrating object detection and classification stages for improved accuracy and robustness.

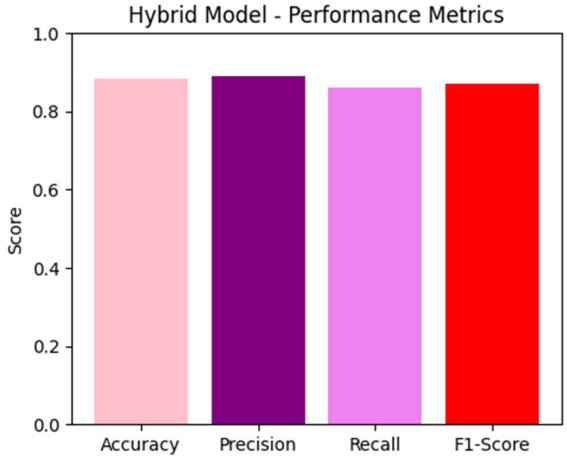


Figure 7. Hybrid model metrics showcases the combined performance of YOLOv8 and CNN.

Figure 8 shows the bar chart comparing inference/processing times and illustrates the efficiency of YOLOv8 and the hybrid model in processing images. YOLOv8's real-time capability is highlighted with an average inference time of 10.8 milliseconds per image, making it suitable for applications requiring fast processing. The hybrid model, while slightly slower at 12.3 milliseconds, still operates within practical limits for real-time

usage. This comparison emphasizes the trade-off between processing speed and enhanced detection/classification accuracy provided by the hybrid approach. Figure 9 shows the overall scores for YOLOv8, CNN, and hybrid YOLOv8-CNN in terms of accuracy, precision, recall, and F1 score, and Figure 10 shows the detection of hybrid YOLOv8-CNN.

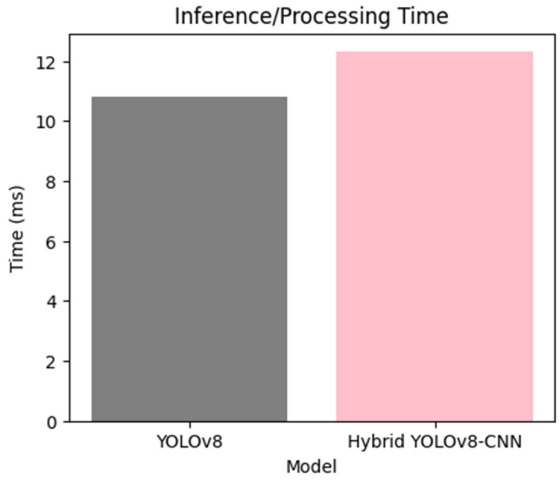


Figure 8. Inference/processing times illustrate the efficiency of YOLOv8 and the hybrid model in processing images.

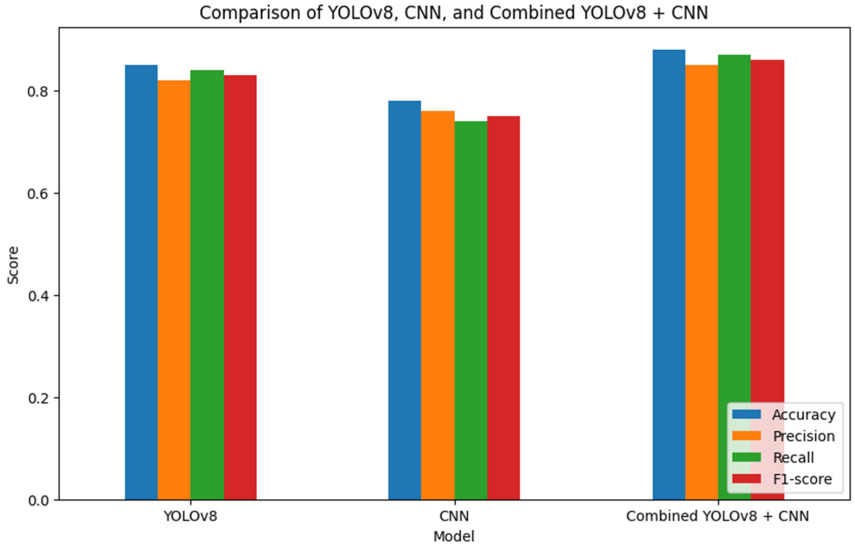


Figure 9. Comparison of YOLOv8, CNN, and hybrid YOLOv8-CNN.

The hybrid YOLOv8-CNN model provides a promising approach to address the limitations of standalone models by effectively combining YOLO's detection strength with CNN's classification accuracy. This approach is particularly valuable in waste management systems that require accurate classification under varying conditions. Future research may explore expanding this hybrid approach to other domains, enhancing its adaptability and performance in diverse, complex environments. Beyond waste management, the proposed hybrid model has potential applications in diverse fields requiring precise object detection and classification. In medical imaging, for example, it could aid in identifying specific cells or anomalies in scans. Autonomous vehicles could utilize the model for real-time detection of pedestrians, vehicles, and other objects on the road. Retail sectors may find value in deploying this model for product recognition, improving inventory tracking, and customer experience in smart retail setups. These applications highlight the model's versatility and importance in real-world deployment.

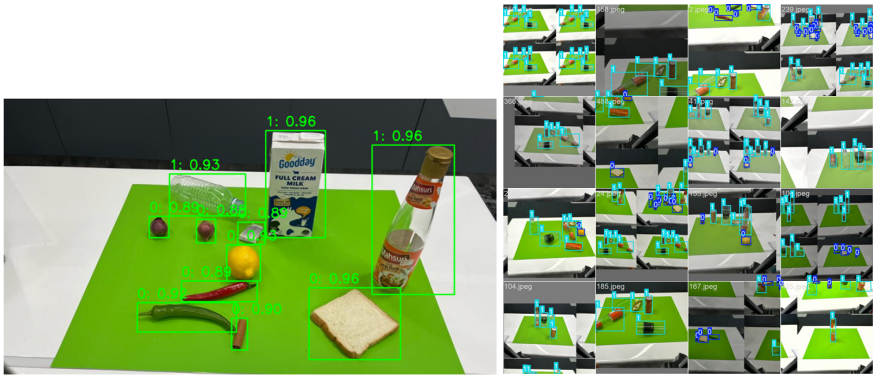


Figure 10. The detection for hybrid YOLOv8-CNN.

A comparison with other hybrid architectures, such as EfficientNet-YOLO and ResNet-CNN models, further emphasizes the choice of YOLOv8-CNN in this work. While EfficientNet and ResNet offer strong performance in classification, their integration with YOLO models has limitations in processing speed and detection accuracy for real-time applications. The YOLOv8-CNN architecture provides a more balanced solution for applications requiring real-time classification without compromising accuracy, especially in the context of object detection for compost and non-compost items.

5. Conclusions

The hybrid model, which combines YOLOv8 and Convolutional Neural Networks (CNNs), can detect and classifying objects. The hybrid model performs more accurately than both the YOLOv8 and CNN models alone. It is especially suitable for the segregation system that separates compostable and non-compostable materials. This integrated approach combines the feature extraction capabilities of standard CNN architecture with the enhanced detection skills of YOLOv8, which is well-known for its real-time object identification performance and accuracy. The method efficiently detects and precisely determines the location of various objects in an image by employing YOLOv8, thus differentiating between compostable and non-compostable materials. The model's capacity to accurately categorize these identified objects is guaranteed by incorporating a CNN, which relies on acquired features. The integration of YOLOv8's fast detection and CNN's deep feature extraction enhances the reliability of the segregation process and reduces the occurrence of misclassification. The model can be further studied for practical applications in which the situations may vary, as it is resistant to changes in object appearance, lighting conditions, and backdrop complications. The effectiveness of YOLOv8 ensures that the automated garbage sorting systems can operate in real-time, which is crucial for their practical deployment. In addition, before the model deployment, further enhancement, especially for the YOLO part, including the application of the latest version of YOLO, may be conducted to increase the detection capability.

Author Contributions: Conceptualization, S.'A.S. and M.I.S.; methodology, A.A.M., S.'A.S., H.S. and M.I.S.; software, A.A.M.; validation, A.A.M., S.'A.S., H.S. and M.I.S.; formal analysis, A.A.M., S.'A.S., H.S. and M.I.S.; investigation, A.A.M. and S.'A.S.; resources, S.'A.S., H.S. and M.I.S.; data curation, S.'A.S., H.S. and M.I.S.; writing—original draft preparation, A.A.M. and S.'A.S.; writing—review and editing, S.'A.S., H.S. and M.I.S.; visualization, A.A.M., S.'A.S., H.S. and M.I.S.; supervision, S.'A.S., H.S. and M.I.S.; project administration, H.S. and M.I.S.; funding acquisition, S.'A.S., H.S. and M.I.S. All authors have read and agreed to the published version of the manuscript.

Funding: 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.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data of this article are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767.
2. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934.
3. Jocher, G. Explore Ultralytics YOLOv8. 2023. Available online: https://docs.ultralytics.com/models/yolov8/#overview (accessed on 19 January 2025).
4. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. In *European Conference on Computer Vision*; Springer: Cham, Switzerland, 2016; pp. 21–37.
5. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollar, P. Focal Loss for Dense Object Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *42*, 318–327. [CrossRef] [PubMed]
6. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
7. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv* **2017**, arXiv:1704.04861.
8. Tan, M.; Le, Q. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; pp. 6105–6114.
9. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In *Advances in Neural Information Processing Systems*; The MIT Press: Cambridge, MA, USA, 2015; Volume 28.
10. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
11. Hussain, M. YOLO-v1 to YOLO-v8, the Rise of YOLO and Its Complementary Nature toward Digital Manufacturing and Industrial Defect Detection. *Machines* **2023**, *11*, 677. [CrossRef]
12. Liu, Z.; Wu, Z.; Dong, Y. YOLO-SE: Improved YOLOv8 for Remote Sensing Object Detection and Recognition. *Appl. Sci.* **2023**, *13*, 12977. [CrossRef]
13. Xu, X.; Liu, X.; Cheng, H. A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *arXiv* **2023**, arXiv:2304.00501.
14. Zhang, C.; Qiao, Y.; Wang, X. Efficient YOLO: An Improved YOLOv8 for Real-Time Object Detection. *IEEE Access* **2023**, *11*, 78965–78978.
15. Brown, J.; Smith, A.; Green, T. Deep Learning Approaches for Waste Classification Using YOLO Models. *J. Environ. Inform.* **2023**, *40*, 456–468.
16. Karna, N.; Putra, M.A.P.; Rachmawati, S.; Abisado, M.; Sampedro, G. Toward accurate fused deposition modeling 3D printer fault detection using improved yolov8 with hy perparameter optimization. *IEEE Access* **2023**, *1*, 1.
17. Chow, L.S.; Tang, G.S.; Solihin, M.I.; Gowdh, N.M.; Ramli, N.; Rahmat, K. Quantitative and Qualitative Analysis of 18 Deep Convolutional Neural Network (CNN) Models with Transfer Learning to Diagnose COVID-19 on Chest X-Ray (CXR) Images. *SN Comput. Sci.* **2023**, *4*, 141. [PubMed]
18. Hong, T.B.; Saruchi, S.A.; Mustapha, A.A.; Seng JL, L.; Razap AN, A.A.; Halisno, N.; Solihin, M.I.; Izni, N.A. Intelligent Kitchen Waste Composting System via Deep Learning and Internet-of-Things (IoT). *Waste Biomass Valorization* **2023**, *15*, 3133–3146. [CrossRef]
19. Tang, G.S.; Chow, L.S.; Solihin, M.I.; Ramli, N.; Gowdh, N.F.; Rahmat, K. Detection of COVID-19 Using Deep Convolutional Neural Network on Chest X-Ray (CXR) Images. In Proceedings of the 2021 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), Virtual, 12–17 September 2021; pp. 1–6.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.