

Article

Real-Time Waste Detection and Classification Using YOLOv12-Based Deep Learning Model

Mosharof Hossain Dipo ¹, Fahmid Al Farid ², Md. Sifti Al Mahmud ¹, Muntasir Momtaz ¹, Shakila Rahman ^{1,*}, Jia Uddin ³ and Hezerul Abdul Karim ^{2,*}

¹ Department of Computer Science, American International University—Bangladesh, Dhaka 1229, Bangladesh; mosharofhossin561@gmail.com (M.H.D.); muntasirtoqi555@gmail.com (M.M.)

² Faculty of Engineering, Multimedia University, Cyberjaya 63100, Malaysia

³ AI and Big Data Department, Woosong University, Daejeon 34606, Republic of Korea; jia.uddin@wsu.ac.kr

* Correspondence: shakila.rahman@aiub.edu (S.R.); hezerul@mmu.edu.my (H.A.K.)

Abstract: Increased waste volume and limitations of traditional separation methods have made waste management a hot topic in recent years. To enable the recycling process to be optimized and to minimize environmental impact, waste materials must be well detected and classified. Building on this research, the system is an automated waste-detecting system that integrates machine vision and artificial intelligence (AI). It is coupled with advanced convolutional neural networks (CNNs), which are used for data collection, real-time waste detection, and classification of the proposed framework. Images of waste were captured in many different settings and analyzed with a YOLOv12-based model. The system achieves more gain in detecting and categorizing waste types with 73% precision and a mean average precision (mAP) of 78% in 100 epochs. Results indicate that the YOLOv12 model surpasses the current detection algorithms to provide an efficient and scalable solution to waste management challenges.

Keywords: YOLOv12; waste detection; machine vision; automated sorting



Academic Editor: George A. Papakostas

Received: 29 March 2025

Revised: 20 May 2025

Accepted: 2 June 2025

Published: 9 June 2025

Citation: Dipo, M.H.; Farid, F.A.; Mahmud, M.S.A.; Momtaz, M.; Rahman, S.; Uddin, J.; Karim, H.A. Real-Time Waste Detection and Classification Using YOLOv12-Based Deep Learning Model. *Digital* **2025**, *5*, 19. <https://doi.org/10.3390/digital5020019>

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

Waste management and environmental sustainability constitute some of the most pressing issues in the world currently because of rapid urbanization, industrialization, and population growth. Effective waste management is a necessity if one intends to reduce the environmental, social, and economic effects. Some recent advances in artificial intelligence, like deep learning and computer vision, open avenues for the automation of waste detection and classification to make the waste management system more efficient and scalable. An important function of drones in defense and surveillance is real-time detection of objects. With the aid of sophisticated imaging devices and high-definition cameras, drones can scan their surroundings continuously and detect and track both stationary and moving targets. This functionality gives security and military professionals valuable information about their surroundings. To perform this task, drones depend on advanced computer vision systems supported by machine learning algorithms. These systems can interpret images and classify different objects by using patterns and features acquired through training. In addition, training these models means drones can learn to identify new kinds of objects. Techniques that have been effective in agricultural grading of products are now being used in waste management, with deep learning approaches such as CNNs, to enhance waste classification and reduce the environmental footprint [1].

Gaps due to the absence of standard benchmarks and datasets meant for waste detection have been filled by establishing novel databases and review studies, providing the backup for better waste recognition and classification [2]. IoT-empowered systems combined with deep learning architectures offer real-time waste separation and monitoring capabilities, enhancing the separation of biodegradable and non-biodegradable materials [3]. Hazardous wastes, including asbestos, are identified by hyperspectral imaging and short-wave infrared analysis; this method is potentially significant in reducing health hazards and improving recycling processes [4]. Lightweight AI models such as EcoDetect-YOLO demonstrate increased capabilities for in-depth domestic waste detection in complex urban landscapes, thus increasing accuracy and speed for small-object detection [5]. AI image recognition technology is revolutionizing municipal solid waste management and is automating sorting and recycling into a sustainable development model [6]. Feature fusion and artificial neural network techniques lay a digital path for automated sorting in smart cities according to circular economy principles [7]. Vision-based methods tackle waste container detection—doing away with traditional RFID technology while improving cost efficiency and flexibility [8]. IoT-integrated smart systems will use technologies such as LoRa and TensorFlow to enable long-range, energy-efficient waste classification and collection [9]. Optimized CNN architectures integrate waste classification models when deployed with genetic algorithms, utilizing datasets such as TrashNet to achieve superior accuracy [10]. Deep learning models can be successfully designed for construction and demolition waste management, improving waste sorting under complex, uncontrolled conditions [11]. AI and sensor technology combinations can be used for creating smart waste separation systems targeting recyclable plastic waste [12]. AI detectors are advancing further into construction and demolition waste sorting, with a focus on very recyclable materials, such as concrete, bricks, and tiles [13]. Hierarchical neural networks in recycling plants offer real-time waste detection and classification, which solves the problems of poor illumination and overlapping materials [14]. Advanced stereo vision systems integrated with CNNs can accurately and effectively achieve multi-object waste detection for road cleaning robots, increasing efficiency and accuracy [15].

The goal of this research is to develop a robust, modular, and scalable waste detection system in support of global sustainability initiatives. This study indicates the possibilities for transformational impact that AI will have on the work of waste management and lays the ground for anticipated developments in multi-object classification, real-time sorting, and wider integrations of AI-powered waste management systems into other environmental conservation programs. Therefore, through these efforts, we want to provide a practical and beneficial answer to one of the most important challenges of our time. The main contributions of this paper are as follows:

- We use high-resolution and low-resolution images with a Yolo-based deep learning model for better identifying waste.
- The model is designed for accurate classification of different types of waste, which improves detection efficiency.
- The proposed structure was tested under different environmental conditions and compared to the related identity model to assess its efficiency.
- It achieves high precision (73%) and a mean average precision (mAP) of 78% over 100 epochs, improving other existing detection algorithms.
- Research makes the base for AI-operated waste management solutions, which contributes to automation and stability in the sorting and recycling of waste.

The results from this research point to AI-based waste detection systems to transform waste management by automating classification, hence improving recycling efficiency and reducing adverse environmental impact. This paper is organized as follows: Section 2

reviews related works on waste detection technologies, Section 3 describes the methodology, Section 4 presents experimental results and their analysis, while Section 5 concludes with future directions to enhance the AI-based waste management system.

2. Literature Review

Waste detection and classification have emerged as important challenges in modern waste management, IoT, deep learning, and extensive research in data-based solutions. Various studies have introduced innovative functions to improve waste sorting and classification efficiency and accuracy. A remarkable study proposes an IoT-based waste insulation system that integrates several sensors, such as moisture, infrared, and proximity, real-time classification of multiple sensors, and real-time classification of metal waste. The system improves the efficiency of waste management by reducing manual labor, sends notifications when rooms are filled, and uses GPS tracking for custom collections. However, challenges such as sensor accuracy, liability, and implementation costs are a matter of concern regarding mass placement [16]. Deep learning models have quickly been used in waste classification. A study using the Yolov8 model achieved better accuracy (97.7%) in the classification of waste types such as plastic, food waste, battery, and glass. The model uses an anchor-free object detection approach, advanced recovery techniques, and a customized loss function to increase accuracy. This improved the first Yolo versions (Yolov4, Yolov5, and Yolov7) and methods for alternative object detections, such as SSD and faster R-CNN. The real-time treatment capacity of Yolov8 makes it very appropriate for identifying autonomous waste in smart cities [17]. Another approach includes the CNN-based VGG16 model for smart waste separation in the home environment, which achieves 84.67% accuracy. The system integrates Raspberry PIE 400, an Ultrasonic Sensor, a Webcam for Image Acquisition, and a Bin Operation Stage. Preparatory image techniques, such as increasing and changing size, are used to increase the strength of the model. The system classified waste in paper, plastic, and glass categories, although its medium accuracy suggests potential reforms through deep network or hybrid models [18]. A multimodal classification strategy introduced a double-merger model across attitudes by combining image and sound data to increase classification accuracy. The system uses 50 iterations as a spine, integrating Jetson Xavier NX and Raspberry Pie 4B for real-time data collection. This achieved an impressive accuracy of 96.24% by maintaining strength against environmental noise and image corruption. The study emphasized the ability of sound-assisted waste classification to separate visually similar types of waste, a struggle for traditional image-based models [19]. Similarly, another study adopts an IoT framework with ultrasound and moisture sensors associated with Arduino Uno, which facilitates automatic pruning of wet and dried waste by increasing the user's involvement through LED indicators. The system includes servo engines for mechanical sorting and monitoring in real time, improving the recycling process. The primary limitation of this approach is its dependence on the dominant waste categories, which limits the compatibility of new types of waste materials [20]. Meanwhile, another study introduced Extent, which is a two-stream transformer-based model designed for the classification of e-waste. By taking advantage of data medical image transformers (DEITs), the model effectively classified e-waste categories such as mobile phones, laptops, keyboards, and microwaves with 96% accuracy. The lightweight architecture of the model, requiring less than one million parameters, demonstrated efficiency in the resource-constrained environment, making it suitable for edge calculation applications [21]. Further research discovered the mobile-based waste classification model, which introduces GMC MobileetV3, which improved the classification performance up to 3.6% over the traditional Mobilentv3. The model includes a Contextual Block Attention Module (CBAM) for better spatial convenience extraction and

a Mish activation feature for a better shield stream. It classifies sorted waste with 96.55% accuracy in kitchens, recyclable, dangerous, and other categories, while the calculation costs are significantly reduced, making it ideal for built-in waste sorting systems [22]. Similarly, another CNN-based approach used a sequential model to classify waste into six categories, adapting the performance with Adam adaptation to achieve high verification accuracy. The architecture of the model consisted of several conv2D layers with batch normalization, maximum pool, dropout, and dense layers. Regardless of the efficiency, the model's dependence on a relatively small dataset (2467 images) limits its generalization capacity and requires large and more different datasets for real-world adequacy [23]. The computer Vision Model has also been utilized to increase the classification of waste. A study that integrates Mobilentv2, Yolov5, a CNN, and Reanet-152 demonstrated high classification accuracy, with MobilentV2 achieving the highest of 97%. The dataset was prepared to increase the amount of normalization, including image standardization and improvement. The study showed that Reset-152 preserved the exact details; it lags in speed compared to Mobilentv2 and Yolov5, making it less suitable for real-time applications [24]. Another research effort focused on plastic waste classification using transfer learning with ResNeXt, DenseNet, and MobileNet_v2, attaining an accuracy of 87.44%. The study utilized the WaDaBa dataset, containing 4000 images of plastic waste categorized into Polyethylene Terephthalate (PETE), High-Density Polyethylene (HDPE), Polypropylene (PP), Polystyrene (PS), and others. Among the models, ResNeXt achieved the highest classification performance, demonstrating the efficacy of residual architectures in plastic waste identification [25]. A recent study introduced RWC-Net, a hybrid model combining DenseNet201 and MobileNet-v2 for effective recyclable waste classification. The model was trained on the TrashNet dataset (2527 images) and achieved 95.01% accuracy. By leveraging feature reuse and lightweight computation, RWC-Net outperformed state-of-the-art models like AlexNet, ResNet, and DenseNet121. Class-specific F1 scores showed high precision for cardboard (97.24%), glass (96.18%), and plastic (93.67%), though performance on litter classification remained relatively lower (88.55%). The authors emphasized the need for further improvements in dataset diversity and the integration of bounding box annotations to enhance practical applications in automated waste management [26]. The progress in IoT-based frameworks, deep learning, and multimodal classification methods has improved waste identification and classification accuracy. However, challenges remain in dataset boundaries, model scalability, and real-world adaptability. Future research should focus on improving AI-operated automation, incorporating self-determined algorithms and integrating additional sensory methods for more extensive waste management solutions.

3. Methodology

This section details the implementation of the proposed waste detection method using a YOLOv12-based object detection framework.

3.1. Data Acquisition

Data collection plays a crucial role in developing a robust waste detection system. In this study, we gathered image data from multiple online sources, including open-source datasets, social media platforms, and environmental agencies. Additionally, custom video recordings were captured using drone footage to enhance dataset diversity. Frames were extracted from these videos to generate a collection of images representing various waste types in different environments. The dataset includes different categories of waste, such as plastic, metal, paper, glass, organic waste, and mixed waste shown in Figure 1.

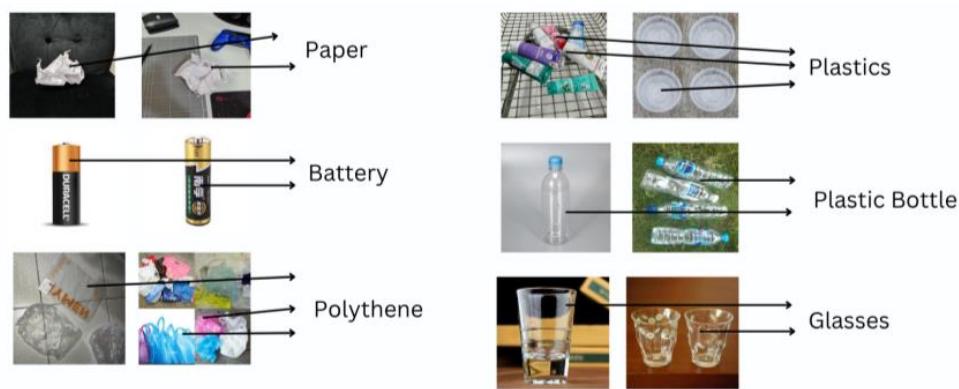


Figure 1. Sample dataset.

3.2. Image Pre-Processing

Data augmentation is a crucial technique that enhances dataset diversity and variability, improving the robustness and accuracy of machine learning models. In this study, augmentation methods were applied to selected images to ensure a more representative and realistic dataset for waste detection. During data collection using a phone camera, some images were captured in low-light conditions, reducing their visibility. To address this, in the data augmentation process, the brightness of each image was altered by randomly increasing or decreasing it by up to 25%. The modification was effective in making overexposed areas brighter and increasing contrast in images, which greatly benefited both image detection and training. To introduce variations in object orientation and enhance the model's generalization capability, a random rotation augmentation technique was applied. Images were randomly rotated between -15 and $+15$ degrees, simulating real-world scenarios where waste items may appear in different orientations. This augmentation helped train the YOLOv12 model to recognize waste items accurately, even when rotated, improving its detection performance in dynamic environments. To further enhance dataset diversity, additional augmentation techniques such as horizontal flipping and scaling were applied. Random horizontal flips allowed the model to learn symmetrical representations of waste objects, while scaling ensured that objects of varying sizes were well-represented in the dataset. These techniques helped mitigate dataset biases and improved the overall robustness of the model. After applying the augmentation techniques, the newly generated images were merged with the original raw dataset to form a more diverse and representative collection. This custom dataset, consisting of both real and augmented images, provided a broader range of variations, significantly improving the model's generalization capability. By incorporating augmented data with different brightness levels, orientations, and sizes, the dataset effectively represented real-world waste detection scenarios. This comprehensive dataset was then split into training (75%), validation (15%), and testing (15%) subsets, ensuring an optimal balance for model learning and evaluation. This approach not only facilitated better training but also improved model performance in real-time waste detection applications. The enhanced dataset allowed for a more accurate, reliable, and scalable waste detection framework, enabling efficient waste management and environmental monitoring.

3.3. Image Resizing and Labeling

Resizing an image means giving all the images the same shape. Thus, the custom dataset with all images was resized to 640×640 pixels. After that, the resized images were labeled according to the seven classes shown in Table 1. The Roboflow annotation operation was applied for labeling data into multiple classes. In Figure 2, we show some sample labeled images for each class. Here, in the first column, the input images are

shown, then in the second column, the class with a bounded box is shown, and lastly, the bounded box layer for each class is shown with different colors in the last column. The class ID in a bounding box is the name of the obstacle. For example, in this study, obstacle labels include Polythene, Battery, Plastic Bottle, HDPE Plastic, and Other Obstacles, as Figure 3 demonstrates. In this case, a bounding box surrounds one obstacle, and the class ID attached to it describes its category.

Table 1. Total images of the dataset.

Types of Obstacles	Quantity Within the Dataset	Total Data
Plastic	4171	
Metal (Battery)	1197	
Paper	1835	
Glass	91	
Organic Waste	112	
Medical Waste	574	7980



Figure 2. (a) Sample data augmentation case 1; (b) sample data augmentation case 2.

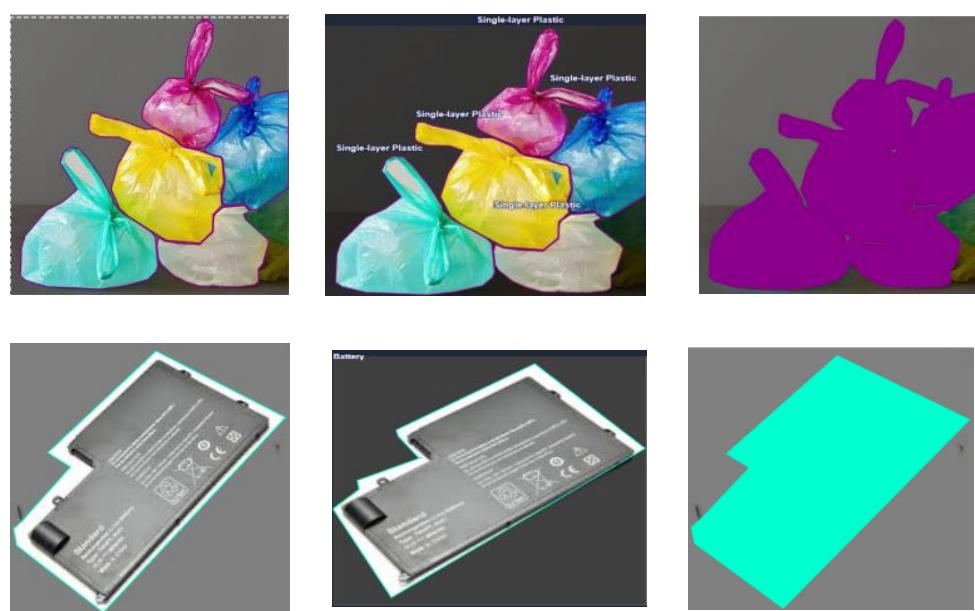


Figure 3. Cont.

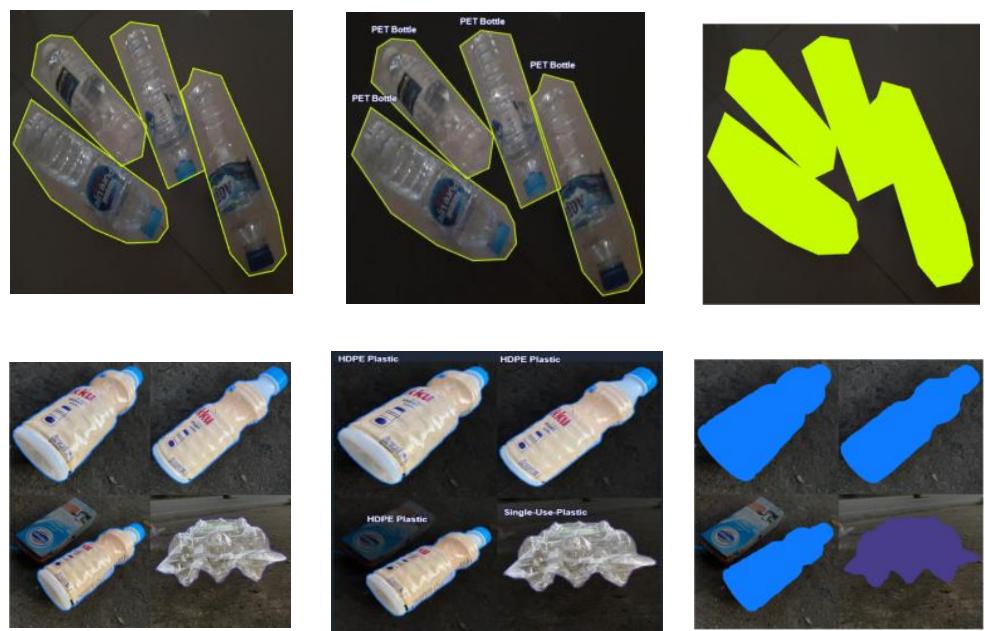


Figure 3. Sample labeled data.

Figure 4 shows the whole sequence of pre-processing steps. After these steps were completed, the processed dataset was divided so that 75% was used for training, 15% for testing, and 15% for validation.

Data Pre-processing Steps

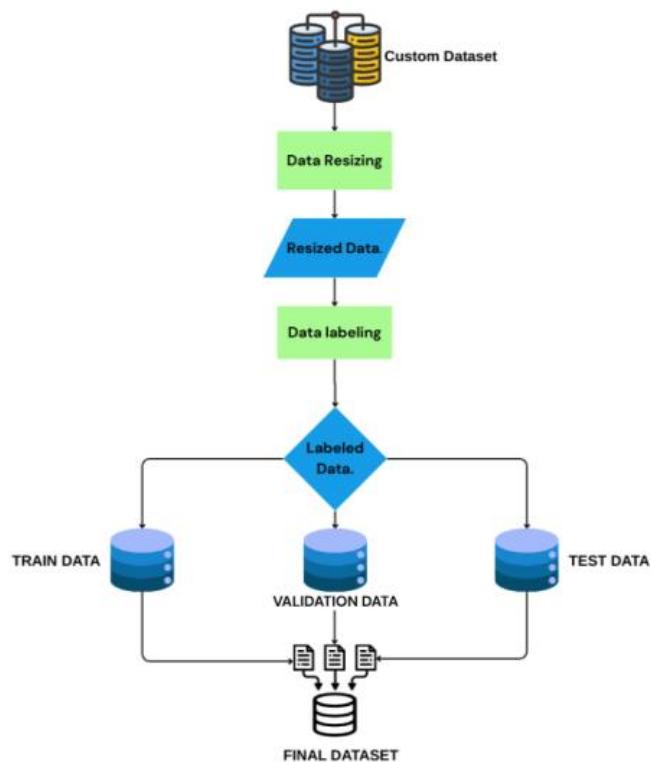


Figure 4. Data pre-processing steps.

3.4. Proposed Obstacle Detection Framework

Figure 5 shows the framework of the suggested Yolov12-based waste identification model, which is carefully designed for effective, real-time, and scalable identification of

various pixel-level waste materials. The architecture benefits from a three-stage processing of pipeline extraction, functional association, and object detection, promoting calculation efficiency without compromising accuracy. Constructed on advanced components such as the region's attention, R-ELANE, and blinking, the system has been adapted to distribution in resources such as mobile and built-in systems.

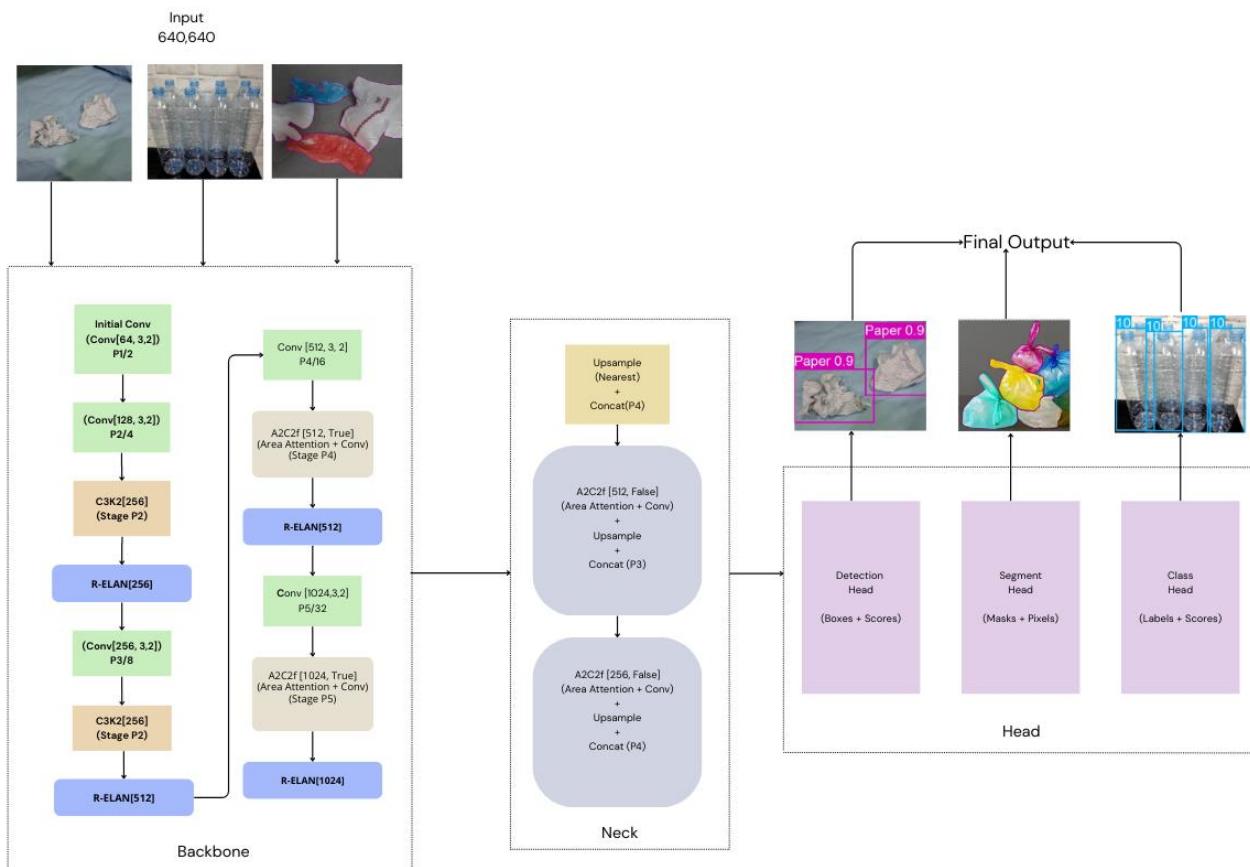


Figure 5. Architecture model.

The spine of the model includes several fixed layers, C3K2 blocks, and SPPF modules to gradually extract hidden features from 640×640 entrance images. These layers reduce spatial dimensions by enriching function maps with meaningful representation. The use of the R-ELE (REST Effective Layer Aggregation Network) introduces residual shortcuts with a scaling factor (standard: 0.01), enabling stable gradient flows and better optimization. Unlike pre-elected architecture, R-ELEN gradually avoids gradient blocking and convergence problems by changing the convenience through a streamlined bottleneck structure. It provides a stable and compact spinal cord, especially favorable in large-scale training scenarios.

In the neck of the model, the architecture integrates up sampling streams, convolution operations, and attention modules (A OFF). The focus of the region holds each function map in simple sections, attracting local attention in each region—it reduces calculation complexity from $O(n \text{ of } n)$ to $O(n)$, and preserves a wide receptive area. This method ensures spatial awareness and relevant understanding of the object boundaries without overhead for global self-attention. In addition, flash attention is faster at this stage by adapting memory access during attention calculation, which results in a significant boost. The MLP expansion ratio is adjusted to 1.2 (under traditional 4.0), causing network loans to increase quickly. These enrichments help to refine low-level functions and high-level functions before detection.

Detection heads are responsible for producing final predictions, including delimitation coordinates, trust points, and waste category labels. Yolov12 eliminates status coding, which depends on the spatial properties learned through the interconnection layers, which simplifies the architecture while maintaining the performance for detection. The last layer integrates features from different parameters to provide accurate detection in several classes. At this stage, a single R-Ellen block is used, which, unlike the three stacked layers of previous Yolo versions, ensures mild calculations and simple convergence. This layout enables real-time detection of waste types such as plastic, glass, batteries, and organic waste under different environmental conditions.

The images taken with smartphone cameras have been prepared using techniques such as frame stabilization, noise filtration, and contrast improvement. This guarantees visual clarity and suitability for intensive learning analysis. All images are 640×640 and generalized.

The enlarged images are manually annotated with a boundary box and class mark. Comprehensive data text (e.g., glitter modulation, rotation, scaling, and flipping) is used to increase dataset diversity and strength. The final dataset is divided into training (75%), verification (15%), and test (15%). The Yolov12 model is trained using this dataset, an adaptation of detection and classification through monitored learning.

The trained model is deployed on a laptop equipped with an NVIDIA Tesla T4 GPU (16 GB VRAM). The equipment used for training and evaluation was sourced from Dell Inc., Round Rock, Texas, United States. The model detects and the location of waste articles in each framework provides an immediate response to sorting autonomous waste or classification functions. Efficient use of calculation and memory ensures that the system is operated evenly in a living environment, from urban roads to industrial waste sites.

4. Experimental Results

This section presents an experimental assessment of the proposed YOLOv12-based waste identification structure. The performance of YOLOv12 is compared to other Yolo variants, including YOLOv8, YOLOv9, YOLOv10, and YOLOv11, using different evaluation matrices. The results are analyzed based on training age, accuracy, recall, and average precision (MAP).

4.1. Hyperparameters

Yolov hyperparameters play an important role in determining the performance and efficiency of 12 models. The selected hypermeter (as listed in Table 2) was fine-tuned through several relapses to ensure optimal training. Batch size: To balance calculation efficiency and model generalization, a batch size of 16 was selected. A small batch size can slow the workflow, while a large one can require extensive memory. The model was trained for 100 epochs to allow enough learning to prevent overfitting. Stochastic gradient descent (SGD) optimizers were used due to their efficiency in handling massive datasets and maintaining stable convergence. The cocoa pre-influencing model was used to move knowledge from a pre-trained model, which accelerates convergence and improves the accuracy of the detection. A learning speed of 0.01 was determined to control the weight update of the model during training, ensuring effective teaching without rigid parameter updates. Over time, a weight loss of 0.0005 was used to prevent overfitting by reducing the size of model parameters. An initial limitation mechanism with a patience value of 100 was used to prevent training if no improvements were seen for more than 100 epochs in a row.

Future research could entail testing a variety of patience intervals to improve the results.

Table 2. Parameters used for the YOLOv12-based object detection model.

Parameters	Value
Batch size	16
Number of epochs	100
Optimizer	SGD
Pre-trained	COCO model
Learning rate	0.01
Weight decay	0.0005
Patience	100

4.2. Model Evaluation

The model assessment performed promisingly in YOLOv12 models that were trained for the proposed waste identification system. Table 3 presents model parameters. The evaluation was performed in a Python 3.10 environment with CUDA 12.0 and NVIDIA-SMI 525.85.12, using a Tesla T4 GPU with 16 GB of graphics memory and 16.0 GB of RAM, sourced from Dell Inc., Round Rock, TX, USA. The trained Yolov12 model, which includes 159 layers and 2,559,848 parameters, performs effective calculations and achieves a GFLOPS value of 6.3. The evaluation included several matrices to assess the effectiveness of the model in waste identification, achieving values of 28.7. The evaluation process included various matrices to assess the model's efficiency in detecting obstacles.

Table 3. Parameters of the evaluation model.

Parameters	Value
Model layers	159
Model parameters	2,559,848
Gradients	2,559,848
GFLOPs	6.3

4.3. Analysis of Results

Comparison of the performance of different Yolo models (Yolov8, Yolov9, Yolov10, Yolov11, and Yolov12) reflects their related strengths, weaknesses, and trends in 50 and 100 training epochs. We have compared them with our own by using a waste detection dataset, and the measures used are F1 score, mAP@0.5, and confusion matrix performance, as presented in Table 4.

Table 4. Testing performance of YOLOv11, YOLOv8, YOLOv9, YOLOv10, YOLOv12, and RF-DETR.

Model	Epoch	Class	Trainable Parameters	F1 Score	mAP@0.5
Proposed YOLOv12	50	All	25.5 M	0.72	0.75
Proposed YOLOv12	100	All	25.5 M	0.75	0.78
YOLOv8 [27]	50	All	25.9 M	0.72	0.72
YOLOv8 [27]	100	All	25.9 M	0.71	0.73
YOLOv9 [28]	50	All	25.3 M	0.69	0.71
YOLOv9 [28]	100	All	25.3 M	0.73	0.75
YOLOv10 [29]	50	All	2.7 M	0.72	0.74
YOLOv10 [29]	100	All	2.7 M	0.74	0.74
YOLOv11 [30]	50	All	27 M	0.74	0.76
YOLOv11 [30]	100	All	27 M	0.75	0.76
RF-DETR	100	All	29 M	0.74	0.78

Yolov12 performed better, with stable training and even losses in loss values and a conversion accuracy of 100 epochs (Figure 6). It recorded the best F1 score of 0.75 and mAP@0.5 of 0.78 and revealed good learning ability and generalization. The precision of 70.73 and recall of 70.77 show a well-set model corresponding to different waste classes. Training loss over time decreased, usually from ~5.05 in epoch 1 to ~1.89 in epoch 100—assessed with initial limitations (patience = 100)—which shows the ongoing learning. Yolov11, the largest model (27.0 m parameter), performed equally to Yolov12, with an F1 score = 0.75 and mAP@0.5 = 0.76 (Figure 7). However, it failed to detect small obstacles sometimes. Further parameters for the model did not receive a predictable accuracy benefit, and its “low generalization capacity” limited performance benefits. The loss of training continued to decrease, but according to the confusion matrix, sometimes small and visually distracting objects were incorrectly identified or ignored. Confusion matrix analysis (Figure 8) reveals good recognition of general classes such as glass bottles and plastic, although the detection rate of extremely short or longer structures that spray was somewhat lower.

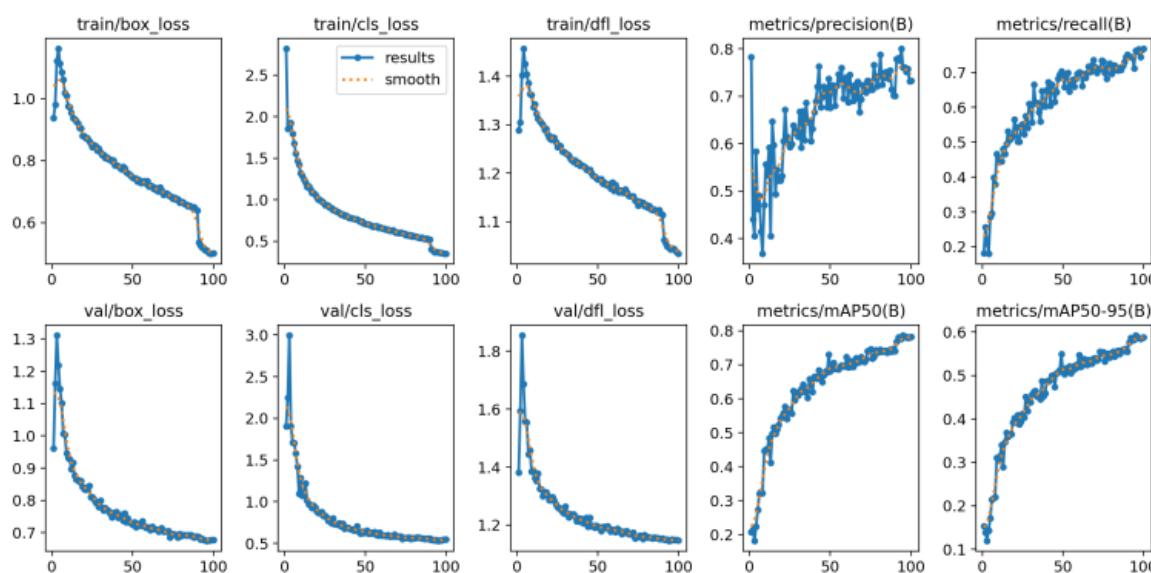


Figure 6. YOLOv12-based training graph with 100 epochs.

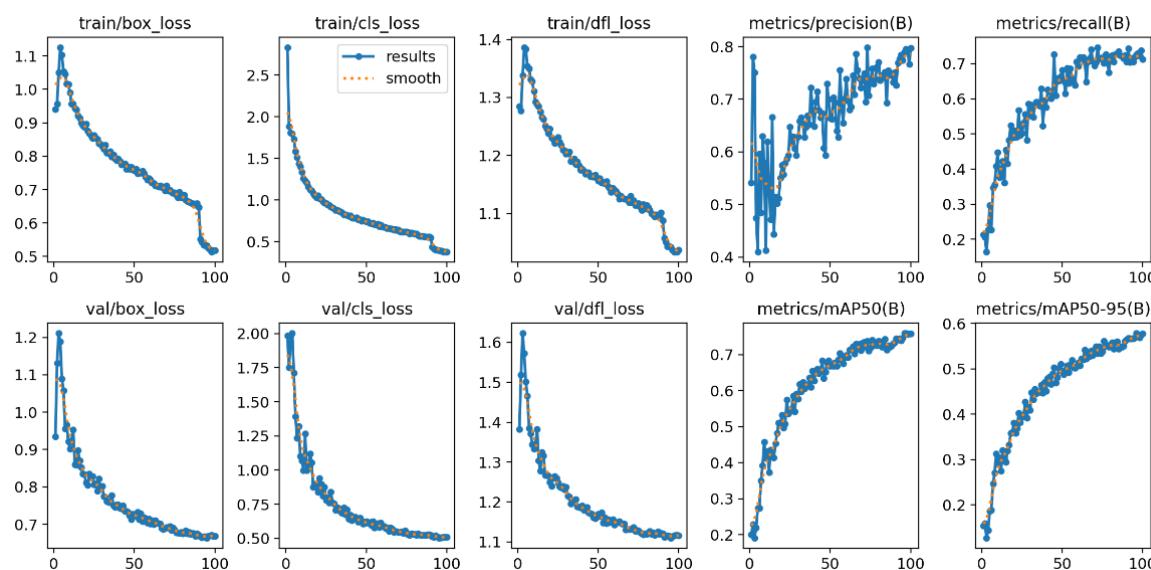


Figure 7. YOLOv11-based training graph with 100 epochs.

Yolov10 performed well continuously, with increasing accuracy, and values were remembered. The training deficit decreased over time, indicating effective learning. However, the model showed signs of plate work in performance, which means that further improvements would require modifications in hyperparameters or datasets. The confusion matrix showed that YOLOv10 sometimes confused objects with similar forms. YOLOv8 had high accuracy and recall, with stable training and confirmation decline. Despite the strong performance, some misclassifications occurred, especially with small or condensed objects. The confusion matrix indicated that Yolov8 effectively detected obstacles but sometimes confused them with background elements. YOLOv12 demonstrated strong learning skills, with mAP and F1 score values. The model performed well overall, and while general accuracy was promising, low abort rates suggested challenges in processing specific object categories. Comparative test view of the proposed YOLOv12 model with other YOLO versions and RF-DETR is condensed in Table 4. The performance was evaluated using our customized dataset, where the average precision (mAP@0.5) at the 0.5 intersection of the union was used to evaluate learning ability. High values indicated better model generalization.

Furthermore, the F1 score is calculated from the equation

$$F1score = (2 \times Precision \times Recall) / (Precision + Recall) \quad (1)$$

Overall, Yolov12 is the most accurate model in the accuracy-to-shape trade-off. The studied models have the highest F1 score and mAP@0.5, with a minor complexity (25.5 m parameters) and stable, smooth exercise dynamics. Including attention to the region, small MLP conditions in R-Allen modules, flashing, and architecture are all additional factors to consider. The efficacy of the YOLOv12-based model for real-time waste detection can be attested from the training graphs in Figures 6 and 7 and the test performance summary in Table 4. The balanced design of the model and stable performance qualifies it as the most suitable model to be deployed in smart waste management systems. The results for RF DETR are $F1 = 0.74$ and $mAP@0.5 = 0.78$. While the accuracy of RF DETR is equal to that of YOLOv12, it takes a little longer (≈ 6 ms/image on an NVIDIA T4) to process images compared to YOLO, which performs well for real-time applications (≈ 5 ms/image). However, direct comparison between YOLOv12 and the RF-DETR model introduces additional variables, such as differences in backbone networks, training procedures, and hyperparameter settings, which complicate the interpretation of results. We focused on the YOLO family of models, as they share similar architectural foundations and training methodologies.

4.4. Visualization

A total of 100 epochs were used for training, with each of them consisting of a complete scanning of the training data and modification of parameters according to signals from loss and gradient calculations. A 50-epoch patience early stopping criterion was applied, so as soon as the model did not improve on the validation set for 50 consecutive epochs, training was terminated at the 100th epoch. On our machines, training the network took roughly 2.492 h, but the time may change according to the computing hardware. Once training is finished, Figure 9 shows several randomly chosen obstacle detections representing the highest level of accuracy, while Figure 10 presents more illustrations of the proposed YOLOv12 model's performance in identifying multiple obstacles in one frame.

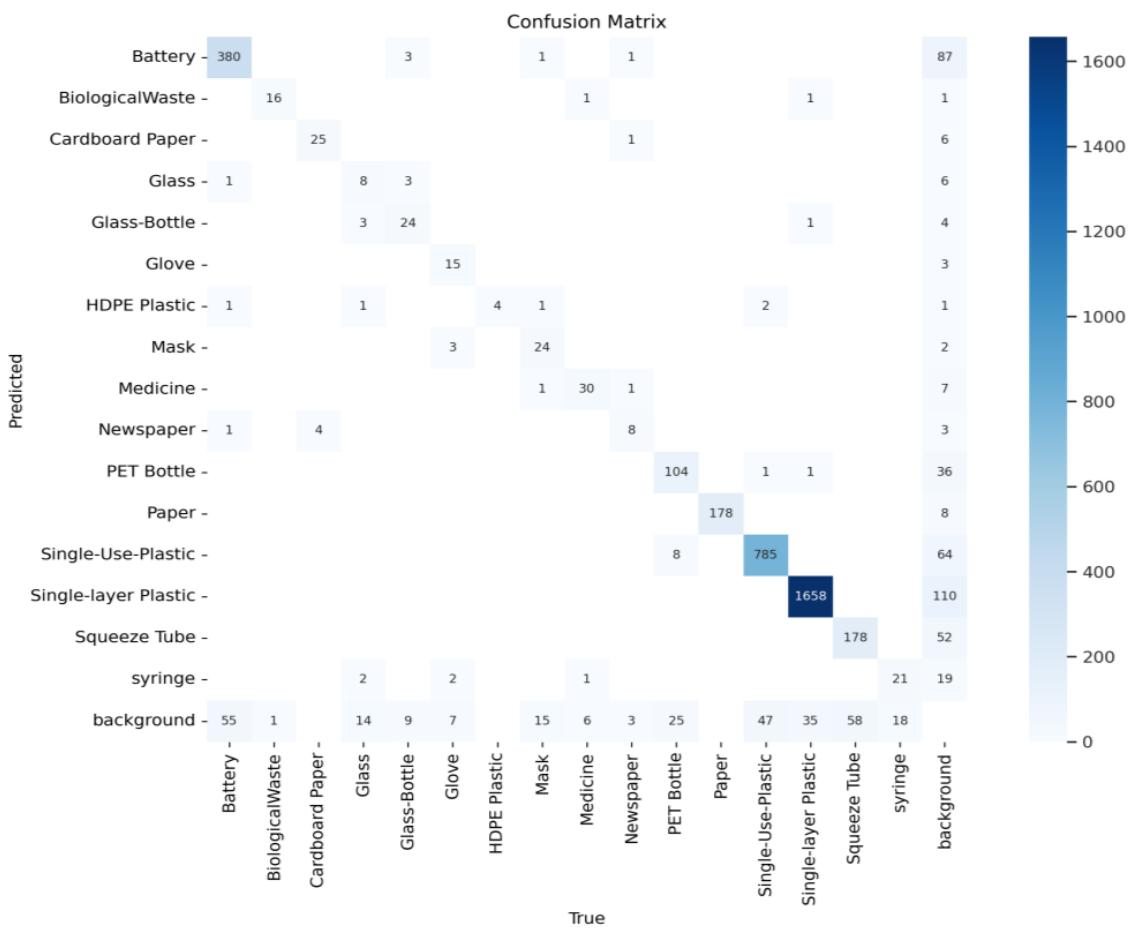


Figure 8. Confusion matrix diagram for 100 epochs.

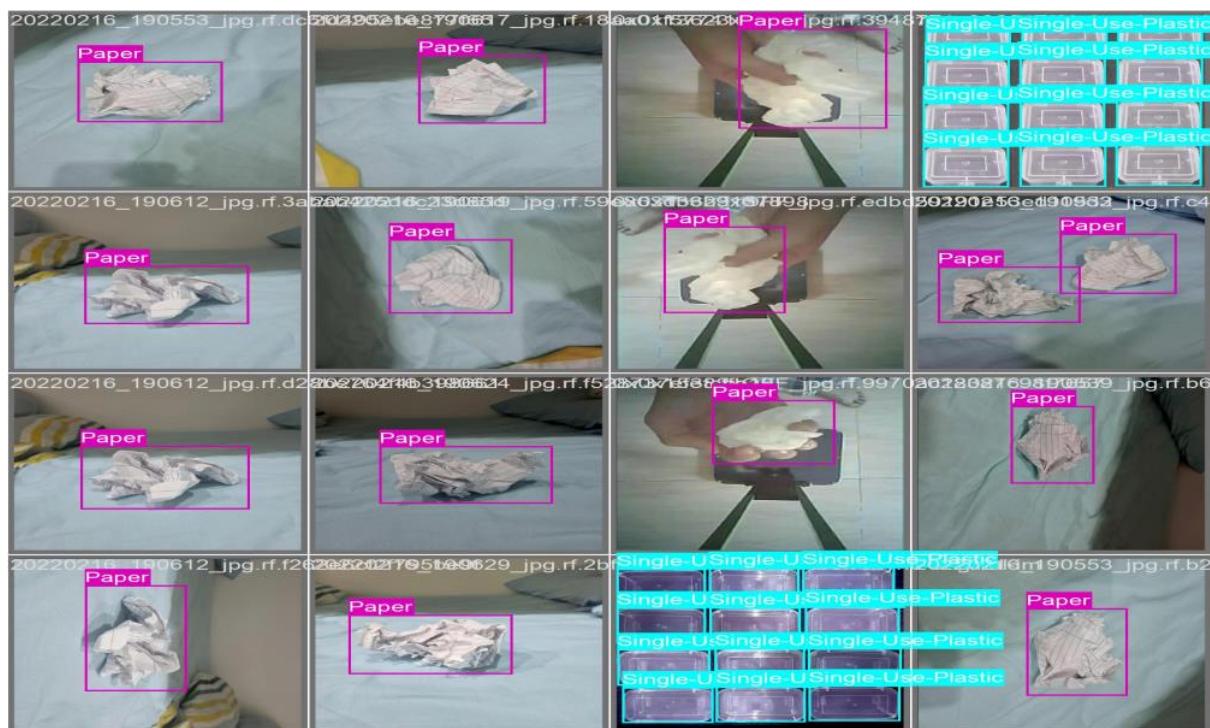


Figure 9. Testing performance for random data.



Figure 10. Examples of multiple detected images in one frame using the YOLOv12-based model.

The picture shows a grid of pictures where the model attempted to detect and classify objects. All detected objects are encircled with a colored bounding box with a label indicating the predicted class (e.g., “glass,” “glass bottle,” “syringe”). Variety of Objects: The pictures contain a variety of objects, including different types of glassware (glasses and bottles) and syringes. The detection quality is inconsistent. There are some objects detected and labeled correctly, while others have lower-accuracy bounding boxes or even incorrect labels. The model seems to perform quite well at detecting and classifying glass items and glass bottles. The bounding boxes are accurately aligned with the objects. There are variations in the shape of glassware detected, including drinking glasses, jars, and bottles of different sizes and shapes. The model also detects syringes in some images. The bounding boxes on the syringes are less precise compared to the detections of glassware. The orientation and shape of the syringes may be the cause of the problem in the detection. Some images contain more than one object (more than one glass or more than one syringe). The model appears to recognize more than one object in one image, which is a good thing. The images present changes in lighting and background. The model appears to handle these variations somewhat, but they can impact the detection accuracy in some cases. The image quality is not uniform. Some are sharper, while the others are blurrier or less resolved. Image quality may affect the performance of object detection models. Potential Improvements and Lessons Learned: The bounding precision of bounding boxes may be improved, particularly for objects like syringes, to better fit the borders of the object. The accuracy of labels of objects may be better tested and improved. Using data augmentation techniques while training can make the model invariant to variations in image quality, lighting, and background. Changing the model architecture or adjusting the

complexity of the current model can improve performance. Having a rich training set with a wide range of object variations and scenarios can make the model more able to generalize. In general, Figure 10 visually depicts the way the model's performance is evaluated on random input examples. The model is promising regarding glass object detection and classification and syringe detection and classification but could improve in bounding box quality, label correctness, and resilience against differences in image and object quality. On the other hand, Figure 11 shows the model accuracy based on different variations of YOLO models, where YOLOv12 performs better in all cases.



Figure 11. Sample detected images for YOLOv12, YOLOv11, and YOLOv10 models.

5. Conclusions

This research presents an advanced identification and classification system for real-time waste that benefits from YOLOv12 deep learning models. This study emphasized how the integration of AI with waste management increases the efficiency and accuracy of sorting by distinguishing between recyclable and non-recyclable waste. By using image datasets, advanced processing techniques, and model optimization, the proposed system achieves an impressive precision of 73% and an average precision (mAP) of 78%. Comparative analysis suggests that the Yolov12 model improves other YOLO variants and strengthens its suitability for real-world applications. The model effectively classifies waste under different conditions and handles challenges such as obstacles and different light environments, which contributes to more efficient and durable waste management. In addition, the AI-driven classification system will reduce work, increase accuracy, and replace waste management practices to promote environmental ability. Adaptability allows integration into the smart urban infrastructure and improves the efficiency of waste collection. The model's ability to process real-time images makes it suitable for distribution in urban and industrial environments, where waste accumulation is a pressure problem,

and its scalability suggests widespread prevention in waste management policy globally. Despite the promising performance, there are some restrictions, such as datasets focusing on specific types of waste, which can prevent normalization, and extreme weather, which can change the lighting and affect the sensitivity of the model to complex backgrounds. In addition, the calculation requirements face challenges in real-time treatment on low-power or edge equipment, and the system sometimes struggles with the detection of small waste objects. Future work involves integrating datasets with real waste categories, integrating IoT-enabled smart rooms and a robot sorting system for improved automation and using learning to be compatible with dynamic waste classification environments. In addition, adaptation of mild AI architecture can improve light distribution on edges, drones, and mobile applications. By combining computer vision with sensory data, such as thermal imaging, hyperspectral analysis or chemical sensors, the accuracy of identification can be further refined, while the development of a pruning mechanism in real time with robotic weapons or conveyor belt integration can increase automation and efficiency.

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

Funding: This research was funded by Multimedia University, Cyberjaya, Selangor, Malaysia (Grant Number: PostDoc(MMUI/240029)).

Data Availability Statement: The dataset is available online https://github.com/MosharofHossainDipo/yolov12_dataset (accessed on 18 May 2025).

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

References

1. Azadnia, R.; Fouladi, S.; Jahanbakhshi, A. Intelligent detection and waste control of hawthorn fruit based on ripening level using machine vision system and deep learning techniques. *Results Eng.* **2023**, *17*, 100891. [[CrossRef](#)]
2. Majchrowska, S.; Mikołajczyk, A.; Ferlin, M.; Klawikowska, Z.; Plantykov, M.A.; Kwasigroch, A.; Majek, K. Deep learning-based waste detection in natural and urban environments. *Waste Manag.* **2022**, *138*, 274–284. [[CrossRef](#)] [[PubMed](#)]
3. Rahman, W.; Islam, R.; Hasan, A.; Bithi, N.I.; Hasan, M.; Rahman, M.M. Intelligent waste management system using deep learning with IoT. *J. King Saud Univ.—Comput. Inf. Sci.* **2022**, *34*, 2072–2087. [[CrossRef](#)]
4. Bonifazi, G.; Capobianco, G.; Serranti, S.; Trotta, O.; Bellagamba, S.; Malinconico, S.; Paglietti, F. Asbestos detection in construction and demolition waste by different classification methods applied to short-wave infrared hyperspectral images. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2024**, *307*, 123672. [[CrossRef](#)] [[PubMed](#)]
5. Liu, S.; Chen, R.; Ye, M.; Luo, J.; Yang, D.; Dai, M. EcoDetect-YOLO: A Lightweight, High-Generalization Methodology for Real-Time Detection of Domestic Waste Exposure in Intricate Environmental Landscapes. *Sensors* **2024**, *24*, 4666. [[CrossRef](#)]
6. Malik, M.; Sharma, S.; Uddin, M.; Chen, C.L.; Wu, C.M.; Soni, P.; Chaudhary, S. Waste classification for sustainable development using image recognition with deep learning neural network models. *Sustainability* **2022**, *14*, 7222. [[CrossRef](#)]
7. Mohammed, M.A.; Abdulhasan, M.J.; Kumar, N.M.; Abdulkareem, K.H.; Mostafa, S.A.; Maashi, M.S.; Chopra, S.S. Automated waste-sorting and recycling classification using artificial neural network and features fusion: A digital-enabled circular economy vision for smart cities. *Multimed. Tools Appl.* **2023**, *82*, 39617–39632. [[CrossRef](#)]
8. Valente, M.; Silva, H.; Caldeira, J.M.; Soares, V.N.; Gaspar, P.D. Detection of waste containers using computer vision. *Appl. Syst. Innov.* **2019**, *2*, 11. [[CrossRef](#)]
9. Sheng, T.J.; Islam, M.S.; Misran, N.; Baharuddin, M.H.; Arshad, H.; Islam, M.R.; Islam, M.T. An internet of things based smart waste management system using LoRa and tensorflow deep learning model. *IEEE Access* **2020**, *8*, 148793–148811. [[CrossRef](#)]
10. Mao, W.L.; Chen, W.C.; Wang, C.T.; Lin, Y.H. Recycling waste classification using optimized convolutional neural network. *Resour. Conserv. Recycl.* **2021**, *164*, 105132. [[CrossRef](#)]

11. Sirimewan, D.; Bazli, M.; Raman, S.; Mohandes, S.R.; Kineber, A.F.; Arashpour, M. Deep learning-based models for environmental management: Recognizing construction, renovation, and demolition waste in-the-wild. *J. Environ. Manag.* **2024**, *351*, 119908. [[CrossRef](#)] [[PubMed](#)]
12. Dokl, M.; Van Fan, Y.; Vujanović, A.; Pintarič, Z.N.; Aviso, K.B.; Tan, R.R.; Čuček, L. A waste separation system based on sensor technology and deep learning: A simple approach applied to a case study of plastic packaging waste. *J. Clean. Prod.* **2024**, *450*, 141762.
13. Demetriou, D.; Mavromatidis, P.; Robert, P.M.; Papadopoulos, H.; Petrou, M.F.; Nicolaides, D. Real-time construction demolition waste detection using state-of-the-art deep learning methods; single-stage vs two-stage detectors. *Waste Manag.* **2023**, *167*, 194–203. [[CrossRef](#)] [[PubMed](#)]
14. Yudin, D.; Zakharenko, N.; Smetanin, A.; Filonov, R.; Kichik, M.; Kuznetsov, V.; Panov, A. Hierarchical waste detection with weakly supervised segmentation in images from recycling plants. *Eng. Appl. Artif. Intell.* **2024**, *128*, 107542. [[CrossRef](#)]
15. Guo, H.; Chen, L. Multi-object road waste detection and classification based on binocular vision. *J. Eng.* **2024**, *2024*, e12389. [[CrossRef](#)]
16. Chavan, O.; Jaware, V.; Doiphode, D.; Deshmukh, P. IoT-Powered Trash Segregation and Waste Management: An Ingenious Approach to a Sustainable Environment. In Proceedings of the 2024 IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT), Greater Noida, India, 9–10 February 2024; Volume 5, pp. 595–598.
17. Bawankule, R.; Gaikwad, V.; Kulkarni, I.; Kulkarni, S.; Jadhav, A.; Ranjan, N. Visual detection of waste using YOLOv8. In Proceedings of the 2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS), Coimbatore, India, 14–16 June 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 869–873.
18. Rijah, U.L.M.; Abeygunawardhana, P.K. Smart waste segregation for home environment. In Proceedings of the 2023 3rd International Conference on Advanced Research in Computing (ICARC), Belihuloya, Sri Lanka, 23–24 February 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 184–189.
19. Xu, H.; Tang, W.; Li, Z.; Qin, K.; Zou, J. Multimodal dual cross-attention fusion strategy for autonomous garbage classification system. *IEEE Trans. Ind. Inform.* **2024**, *20*, 13319–13329. [[CrossRef](#)]
20. Badoni, P.; Walia, R.; Mehra, R. Enhancing waste separation and management through IoT system. In Proceedings of the 2024 1st International Conference on Innovative Sustainable Technologies for Energy, Mechatronics, and Smart Systems (ISTEMS), Dehradun, India, 26–27 April 2024; pp. 1–6.
21. Islam, N.; Jony, M.M.H.; Hasan, E.; Sutradhar, S.; Rahman, A.; Islam, M.M. Ewastenet: A two-stream data efficient image transformer approach for e-waste classification. In Proceedings of the 2023 IEEE 8th International Conference on Software Engineering and Computer Systems (ICSECS), Penang, Malaysia, 25–27 August 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 435–440.
22. Tian, X.; Shi, L.; Luo, Y.; Zhang, X. Garbage classification algorithm based on improved mobilenetv3. *IEEE Access* **2024**, *12*, 44799–44807. [[CrossRef](#)]
23. Gill, K.S.; Anand, V.; Gupta, R. Garbage Classification Utilizing Effective Convolutional Neural Network. In Proceedings of the 2023 2nd International Conference on Vision Towards Emerging Trends in Communication and Networking Technologies (ViTECoN), Vellore, India, 5–6 May 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–4.
24. Kumar, R.L.; Ramya, R.; Balaji, M.J.; Hari, V.; Malarvizhi, M. Garbage Collection and Segregation using Computer Vision. In Proceedings of the 2024 International Conference on Inventive Computation Technologies (ICICT), Greater Noida, India, 19–20 February 2021; IEEE: Piscataway, NJ, USA, 2024; pp. 1023–1028.
25. Asha, V.; Govindaraj, M.; Kolambkar, M.L.; Mithuna, P.; Prasad, A. Classification of Plastic Waste Products using Deep Learning. In Proceedings of the 2024 1st International Conference on Cognitive, Green and Ubiquitous Computing (IC-CGU), Bhubaneswar, India, 1–2 March 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 1–6.
26. Hossen, M.M.; Majid, M.E.; Kashem, S.B.A.; Khandakar, A.; Nashbat, M.; Ashraf, A.; Chowdhury, M.E. A reliable and robust deep learning model for effective recyclable waste classification. *IEEE Access* **2024**, *12*, 13809–13821. [[CrossRef](#)]
27. Rahman, S.; Rony, J.H.; Uddin, J.; Samad, M.A. Real-Time Obstacle Detection with YOLOv8 in a WSN Using UAV Aerial Photography. *J. Imaging* **2023**, *9*, 216. [[CrossRef](#)]
28. Sirajus, S.; Rahman, S.; Nur, M.; Asif, A.; Harun, M.B.; Uddin, J.I.A. A Deep Learning Model for YOLOv9-based Human Abnormal Activity Detection: Violence and Non-Violence Classification. *IJEEE* **2024**, *20*, 3433.
29. Mao, M.; Lee, A.; Hong, M. Efficient Fabric Classification and Object Detection Using YOLOv10. *Electronics* **2024**, *13*, 3840. [[CrossRef](#)]
30. Navin, N.; Farid, F.A.; Rakin, R.Z.; Tanzim, S.S.; Rahman, M.; Rahman, S.; Uddin, J.; Karim, H.A. Bilingual Sign Language Recognition: A YOLOv11-Based Model for Bangla and English Alphabets. *J. Imaging* **2025**, *11*, 134. [[CrossRef](#)] [[PubMed](#)]

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.