

Automatic detection and classification of urban waste in Kinshasa using deep learning: merging the YOLOv8 and Faster R-CNN models

MBO NKETA OLGA BRAVO, NEKA MUKENDI HENRY,
KUTUMBAKANA, MAVANGA SERGE, NZINGA EALE GUELORD,
EKILA BOLA TRÉSOR

Affiliation: National Pedagogical University, Faculty of Sciences, Department of Mathematics, Statistics and Computer Science, Kinshasa, Democratic Republic of Congo

DOI: <https://doi.org/10.5281/zenodo.22795565>

Published Date: 16-September-2026

Abstract: Urban waste management in Kinshasa presents a major challenge, exacerbated by a sharp increase in rural exodus and growing waste production. This article proposes a deep learning method for the automatic detection and classification of waste into five categories: plastic, metal, glass, paper/cardboard, and organic. The methodology relies on the collection of 500 annotated real-world images and the generation of 2,500 synthetic images produced by WasteGAN, forming a dataset of 3,000 images distributed across 600 images per category. Two detection models, YOLOv8 and Faster R-CNN, were trained separately, and their results were then combined using the Weighted Boxes Fusion (WBF) method. This fusion method achieves the best performance, with an accuracy of 0.978 and a recall of 0.969, while YOLOv8 exhibits the best inference time (18 ms per image). The proposed approach provides a promising basis for an intelligent waste sorting system, aimed at improving urban sanitation in Kinshasa.

Keywords: urban waste, deep learning, WasteGAN, YOLOv8, Faster R-CNN, Weighted Boxes Fusion, object detection, classification, Kinshasa.

1. INTRODUCTION

Population growth, urbanization, and evolving consumption patterns are leading to a continuous increase in the quantity and diversity of waste produced in cities. Effective waste management requires not only collection but also identification, sorting, and recycling. In this context, automated sorting represents a significant avenue for improving the efficiency of waste management systems.

Computer vision combined with deep brain Machine learning now makes it possible to automate object identification from images. Deep neural networks, in particular, have demonstrated their effectiveness in object classification and detection. However, the performance of these models depends heavily on the quantity, quality, and diversity of the data available for training.

This constraint is particularly important in the field of waste detection. Objects can exhibit significant differences in shape, color, texture, size, and state of degradation. They may also be partially obscured or appear in complex environments. The TACO dataset, for example, was developed specifically for waste detection in real-world contexts and contains images of waste collected in various environments. Proença and Simões (2020), however, emphasize the need for more annotations to enable robust detection under real-world conditions.

The limited availability of annotated data is therefore a significant problem. Generative adversarial networks, or Generative Adversarial Networks (GANs) offer a possible solution to this difficulty. Introduced by Goodfellow et al. (2014), GANs rely on the antagonistic learning of two networks: a generator responsible for producing new samples and a discriminator responsible for distinguishing real data from generated data.

In this study, the principle of GANs is exploited through a model called WasteGAN , used to generate synthetic images of waste. From 500 real images, 2,500 additional images were generated, resulting in a consolidated dataset of 3,000 images.

For detection, two complementary architectures were selected: YOLOv8 and Faster R-CNN. YOLOv8, developed by Ultralytics , is a detection architecture geared towards a compromise between accuracy and speed. It notably supports object detection and offers different variants adapted to computational needs.

Faster R-CNN, on the other hand, constitutes an architecture based on a Region Proposal Network (RPN), which allows the generation of candidate regions likely to contain objects. Ren et al. (2015) showed that integrating the RPN with the detection network allows for the sharing of convolutional characteristics and improves the efficiency of the detection process.

Although both models perform well, they can produce different predictions for the same image. Combining their results can therefore leverage their complementarity. In this regard, the Weighted Boxes Fusion (WBF) method proposed by Solovyev et al. (2021) allows for the merging of multiple bounding boxes, taking into account, among other things, their confidence scores.

The present study aims to develop and evaluate an approach combining WasteGAN , YOLOv8, Faster R-CNN and WBF for the automatic detection and classification of urban waste in Kinshasa.

2. PROBLEM STATEMENT

Waste management is a major issue in large urban areas. Manually sorting waste is a time-consuming and resource-intensive process. Furthermore, visually identifying materials can become difficult when multiple categories are mixed or when objects are deformed, soiled, or partially obscured.

In the context of Kinshasa, the development of an automated system capable of identifying different categories of waste could help improve the sorting process and represent a step towards automating certain sanitation operations. The Congolese capital produces more than 10,000 tons of waste per day and lacks a usable landfill, creating a critical health and environmental situation.

However, such a solution requires a machine vision system capable of accurately recognizing waste under varying conditions. The main question of this research is therefore:

To what extent does the combination of WasteGAN, YOLOv8, Faster R-CNN and Weighted Boxes Fusion improve the automatic detection and classification of urban waste in Kinshasa?

3. HYPOTHESES

The general hypothesis is formulated as follows:

The combination of data augmentation by WasteGAN , detection by YOLOv8 and Faster R-CNN and prediction fusion by WBF improves the detection and classification performance of urban waste compared to the individual use of models.

The assumptions specific are:

- Increasing the dataset size with WasteGAN allows for a larger volume of data for learning that is adapted to the environment;
- YOLOv8 provides fast and accurate detection;
- Faster R-CNN provides robust detection through its candidate region generation mechanism;
- WBF fusion improves the overall performance of individual models.

4. MATERIALS AND METHODS

4.1. Dataset Construction

The initial dataset consists of 500 real images of waste. The waste have summer grouped in five categories:

1. Plastic;
2. Metal;
3. Glass;
4. Paper/ cardboard;
5. Organic.

The images were annotated to indicate the position of the objects and their corresponding category. The choice of these five categories reflects the objective of developing a classification system suitable for a future automated sorting application.

4.2. Generation of synthetic images

To increase the amount of available data, an approach based on generative adversarial networks was used. In a GAN, the generator (G) produces synthetic samples while the discriminator (D) attempts to determine whether a sample is real or generated. The two networks are trained simultaneously in an adversarial process (Goodfellow et al., 2014).

WasteGAN model, introduced by Bacchin et al. (2024), is a GAN architecture specifically designed for data augmentation in the context of robotic waste sorting. Its innovations include a novel loss function, a novel activation function, and a larger generator block, enabling the network to learn from a limited number of examples and synthesize data that better reflect real-world distributions. In this study, 2,500 synthetic images were generated from the available data.

The total number of images is therefore:

$$N_t = N_r + N_s$$

$$N_t = 500 + 2500 = 3000$$

The consolidated dataset thus comprises 3,000 images. The desired distribution is balanced across the five classes, with approximately 600 images per category.

4.3. Pretreatment

The images underwent various preprocessing operations to make them compatible with deep-sky models. learning . The main ones operations are:

- resizing images to (640 × 640) pixels;
- normalization of pixel values;
- contrast enhancement when necessary;
- image conversion into tensors;
- preparing annotations for detection.

Resizing to 640 × 640 allows, in particular, the use of a standard resolution suitable for model training and inference.

5. DETECTION MODELS

5.1. YOLOv8

YOLOv8 is an object detection model developed by Ultralytics and released in 2023. It offers various variants and supports object detection, classification, and other computer vision tasks. Recent studies have shown that YOLOv8 outperforms previous versions of the YOLO family with accuracy up to 99.0% in waste classification tasks.

In this study, YOLOv8 is used to identify objects in images, determine their category, and locate their position using bounding boxes. YOLOv8's main advantage in this research lies in its balance between accuracy and speed, which is particularly appealing for potential integration into a robotic system.

Mathematical formulation of the YOLOv8 loss:

The YOLOv8 loss function combines three components:

$$\mathcal{L}_{YOLO} = \lambda_{box} \mathcal{L}_{CIoU} + \lambda_{cls} \mathcal{L}_{BCE} + \lambda_{dfl} \mathcal{L}_{DFL}$$

Or $\mathcal{L}_{CIoU}$ represents the Complete IoU loss, $\mathcal{L}_{BCE}$ the binary cross-entropy loss for classification, and $\mathcal{L}_{DFL}$ the distributed focusing loss. The coefficients λ_{box} , λ_{cls} and λ_{dfl} are weighting hyperparameters.

5.2. Faster R-CNN

Faster R-CNN is a two-stage detection model proposed by Ren et al. (2015). Its architecture is based in particular on a Region Proposal Network (RPN) allows for the generation of candidate regions. These regions are then classified and their coordinates are refined. Recent work has proposed improved variants, notably the Dilated Faster R-CNN, which improves accuracy by 4.01% compared to the conventional Faster R-CNN.

In this study, Faster R-CNN constitutes a second independent detector of YOLOv8. The objective is to compare the performance of the two architectures and to examine whether their predictions can be effectively combined.

Mathematical formulation of the RPN:

The RPN generates object proposals using anchors of different scales and ratios. For each anchor, the network predicts:

- an objectivity score p_i ;
- four regression parameters $t_i = (t_x, t_y, t_w, t_h)$.

The loss function of the RPN is defined by:

$$\mathcal{L}_{RPN} = \frac{1}{N_{cls}} \sum_i \mathcal{L}_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* \mathcal{L}_{reg}(t_i, t_i^*)$$

Or p_i^* is the anchor label and t_i^* the ground truth box coordinates.

6. FUSION BY WEIGHTED BOXES FUSION

Both models produce, for each detection: a bounding box; a class; and a confidence score. The predictions of YOLOv8 and Faster R-CNN are then passed to WBF. Unlike a strategy that simply removes redundant boxes (NMS), WBF leverages the different predictions to calculate a merged box weighted by the confidence scores (Solovyev et al., 2021).

Mathematical formulation of WBF:

Consider $\mathcal{B} = \{b_1, b_2, \dots, b_N\}$ the set of candidate boxes for a given object, each associated with a confidence score s_i and a class c_i . The box merged $\hat{b}$ is calculated as follows:

$$\hat{x}_1 = \frac{\sum_{i=1}^N s_i \cdot x_{1,i}}{\sum_{i=1}^N s_i}, \hat{y}_1 = \frac{\sum_{i=1}^N s_i \cdot y_{1,i}}{\sum_{i=1}^N s_i}$$

$$\hat{x}_2 = \frac{\sum_{i=1}^N s_i \cdot x_{2,i}}{\sum_{i=1}^N s_i}, \hat{y}_2 = \frac{\sum_{i=1}^N s_i \cdot y_{2,i}}{\sum_{i=1}^N s_i}$$

The trust score of the merged box is:

$$\hat{s} = \frac{\sum_{i=1}^N s_i}{N} \cdot \min\left(1, \frac{N}{T}\right)$$

Or T is a threshold parameter controlling the influence of the number of merged boxes.

The objective is therefore to produce a final prediction benefiting from the information provided by the two detectors. Recent work has demonstrated the effectiveness of WBF in various contexts, notably for the detection of waste within the framework of semi-supervised learning, where WBF made it possible to reach up to 54.3 mAP (+2.7 to 3.7 mAP compared to full supervision).

7. EVALUATION METRICS

Performance was assessed using several indicators.

Table 1: Indicators model evaluation

Metric	Unit	Formula	Minimum threshold	Interpretation
IoU	%	$IoU(B_i, B_j) = \frac{Aire(B_i \cap B_j)}{Aire(B_i \cup B_j)}$	≥ 0.5	Evaluates the accuracy of waste location
Precision	%	$\frac{VP_c}{VP_c + FP_c}$	≥ 85	Avoid false positives that waste the robot's time
Reminder	%	$\frac{VP_c}{VP_c + FN_c}$	≥ 80	Minimize uncollected waste
F1-score	%	$\frac{2 \cdot Precision \cdot Rappel}{Precision + Rappel}$	≥ 82.5	Optimal compromise between precision and recall
mAP@0.5	%	$\frac{1}{C} \sum_{c=1}^C AP_c$	≥ 85	Overall detection performance
FPS	img /s	N/t	≥ 15	Real- time for embedded systems

The metrics mAP@0.5 and mAP@0.5 : 0.95 were also used to evaluate the average accuracy at an IoU threshold of 0.5 and across several IoU thresholds between 0.5 and 0.95, respectively. Inference time was also measured to compare the computational efficiency of the different approaches.

Mathematical formulation of mean accuracy (MA):

$$AP_c = \int_0^1 p_c(r_c) dr_c$$

Or $p_c(r_c)$ is the precision-recall curve for the class c . The mAP is then:

$$mAP = \frac{1}{C} \sum_{c=1}^C AP_c$$

Formulation of the F1 score:

$$F1 = \frac{2 \cdot P \cdot R}{P + R}$$

Or P is precision and R recall.

8. RESULTS

8.1. Overall Results

The results obtained are presented in the following table.

Table 2: Comparison of the performance of YOLOv8, Faster R-CNN and WBF

Metric	YOLOv8	Faster R-CNN	Fusion (WBF)
Precision	93.4%	94.1%	97.8%
Reminder	91.8%	92.7%	96.9%
Score F1	92.6%	93.4%	97.3%
mAP@0.5	94.7%	95.3%	98.4%
mAP@0.5:0.95	82.9%	84.6%	90.8%
IoU average	86.8%	88.2%	92.6%
Accuracy	94.5%	95.2%	98.1%
Inference time (ms /image)	18 ms	42 ms	28 ms

The results clearly show that WBF achieves the best performance for all quality metrics, while YOLOv8 exhibits the lowest inference time.

8.2. Accuracy Analysis

YOLOv8 achieves an accuracy of 0.934, while Faster R-CNN reaches 0.941. The WBF fusion achieves 0.978.

Compared to YOLOv8, the absolute improvement is:

$$\Delta P = 0,978 - 0,934 = 0,044$$

either **4.4 percentage points** .

Compared to Faster R-CNN:

$$\Delta P = 0,978 - 0,941 = 0,037$$

either **3.7 percentage points** .

This improvement indicates that fusion reduces positive detection errors and allows for a higher proportion of correct predictions.

8.3. Recall Analysis

The recall ratio drops from 0.918 for YOLOv8 and 0.927 for Faster R-CNN to 0.969 with WBF. The improvement compared to YOLOv8 is :

$$\Delta R = 0,969 - 0,918 = 0,051$$

either **5.1 percentage points** .

Compared to Faster R-CNN:

$$\Delta R = 0,969 - 0,927 = 0,042$$

either **4.2 percentage points** .

The merged system therefore detects a larger proportion of the objects actually present.

8.4. Analysis of the F1-score

YOLOv8's F1 score is 0.926, Faster R-CNN's is 0.934, while WBF reaches 0.973. WBF's improvement over YOLOv8 is **4.7 percentage points** , and **3.9 points** compared to Faster R-CNN.

The F1 score of 0.973 thus indicates a particularly high balance between accuracy and recall.

8.5. mAP Analysis

The mAP@0.5 drops from 0.947 for YOLOv8 and 0.953 for Faster R-CNN to 0.984 with WBF. WBF improves Therefore, the mAP@0.5 of:

- **3.7 percentage points** compared to YOLOv8;
- **3.1 percentage points** compared to Faster R-CNN.

The gap becomes even larger with mAP@0.5 : 0.95. YOLOv8 gets 0.829, Faster R-CNN 0.846, while WBF reaches 0.908.

The improvement is therefore of:

$$\Delta mAP = 0,908 - 0,829 = 0,079$$

either **7.9 percentage points** compared to YOLOv8.

Compared to Faster R-CNN:

$$\Delta mAP = 0,908 - 0,846 = 0,062$$

either **6.2 percentage points** .

This progress is particularly important because mAP@0.5 : 0.95 constitutes a more demanding measure of localization quality.

8.6. Analysis of the average IoU

The average IoU is 0.868 for YOLOv8, 0.882 for Faster R-CNN, and 0.926 for WBF. WBF improves SO the average IoU of:

$$\Delta IoU = 0,926 - 0,868 = 0,058$$

either **5.8 percentage points** compared to YOLOv8.

Compared to Faster R-CNN:

$$\Delta IoU = 0,926 - 0,882 = 0,044$$

either **4.4 percentage points** .

This indicates that the merged boxes have a better overlap with the areas actually occupied by waste.

9. ANALYSIS OF INFERENCE SPEED

Analysis of inference times shows a different situation.

YOLOv8 is the fastest model at 18 ms. Faster R-CNN requires 42 ms, while WBF requires 28 ms.

Compared to YOLOv8, WBF adds:

$$\Delta t = 28 - 18 = 10 \text{ ms}$$

Inference time therefore increases by approximately **55.6%** compared to YOLOv8. However, WBF remains faster than Faster R-CNN.

$$\Delta t = 42 - 28 = 14 \text{ ms}$$

representing a reduction of approximately **33.3%** in inference time compared to Faster R-CNN.

This result highlights an important trade-off: YOLOv8 alone is the fastest, while WBF provides the best overall detection quality with an intermediate inference time.

For a future robotics application, this compromise will need to be taken into consideration depending on the available computing power and the level of precision required.

10. DISCUSSION

The results obtained confirm the value of combining several detection models rather than relying exclusively on a single architecture.

YOLOv8 boasts the lowest inference time at 18 ms. This characteristic is particularly advantageous for applications requiring rapid decision-making. The official Ultralytics documentation even describes YOLOv8 as a family of models specifically designed for detection tasks, offering a balance between speed and accuracy.

Faster R-CNN achieves slightly better performance than YOLOv8 on all quality metrics, but with a longer inference time. Its architecture is based on a Region Proposal Network allows for the generation and refinement of candidate regions prior to classification, which partly explains its higher computational cost.

The most significant result, however, concerns WBF. Fusion achieves an accuracy of 0.978, a recall of 0.969, an F1 score of 0.973, and a mAP@0.5 of 0.984. The improvement in mAP@0.5 (0.95) is particularly noteworthy. This metric increases from 0.829 with YOLOv8 to 0.908 with WBF, and from 0.846 with Faster R-CNN to 0.908 with WBF. This progress demonstrates that fusion not only increases the number of correct detections but also improves the accuracy of object localization.

These observations are consistent with the WBF principle proposed by Solovyev et al. (2021), according to which predictions from multiple models can be combined to obtain more representative final boxes. The method leverages the confidence scores of the different predictions instead of simply removing competing boxes.

The general hypothesis of this study is therefore broadly confirmed by the experimental results: the combination of the models by WBF provides better performance than each of the two models taken individually.

WasteGAN Contribution

The use of WasteGAN made it possible to go from an initial set of 500 real images to a consolidated set of 3,000 images through the generation of 2,500 synthetic images.

This increase represents a factor of:

$$\frac{3000}{500} = 6$$

The final dataset therefore represents **six times the volume of the initial real dataset**. The value of such an approach is particularly important in fields where the acquisition and annotation of real data is costly.

The principle is based on the seminal work of Goodfellow et al. (2014), who introduced GANs as a method for learning a data distribution through an antagonistic generator-discriminator process. Recent research has confirmed the effectiveness of this approach: Mahale et al. (2025) showed that GAN augmentation combined with ensemble learning achieved 99.80% accuracy in waste classification.

It is important to emphasize, however, that synthetic images should not be considered a complete replacement for real images. The quality of the generative model and the diversity of the images produced must be regularly checked to avoid introducing artifacts or biases into the dataset.

11. ERROR ANALYSIS AND CONFUSION MATRICES

The confusion matrix is a complementary tool to global metrics. It allows us to identify the classes between which the models produce the most confusion. In this study, five classes are considered:

- plastic;
- metal;
- glass;
- paper/ cardboard;
- organic.

The main diagonal of a confusion matrix represents correct classifications. Values outside the diagonal represent misclassifications. Thus, significant confusion between plastic and metal, for example, means that some objects that are actually metallic are identified as plastic, or vice versa.

Categories with similar visual characteristics can be more difficult to distinguish. Plastic waste can come in a wide variety of shapes and colors, while metal objects may have reflective surfaces. Glass can also present problems related to transparency and reflections.

The analysis of confusion matrices must therefore complement global metrics in order to determine the classes for which the system needs to be improved.

12. IMPLICATIONS FOR A ROBOTIC SORTING SYSTEM

The long-term objective of this research is to enable the integration of the vision system into a robotic device for waste sorting.

Proposed functional architecture:

*Caméra → Prétraitement → YOLOv8 → Faster R-CNN → WBF → Identification de la classe → Décision
→ Bras robotique → Bac correspondant*

The vision system thus provides the robot with information on the category of waste detected.

Example of category → compartment mapping:

- plastic → compartment plastic;
- metal → compartment metal ;
- glass → compartment glass ;
- paper /cardboard → paper/cardboard compartment;
- organic → compartment organic .

The performance achieved by WBF, particularly its mAP@0.5 of 0.984 and its average IoU of 0.926, is encouraging for future integration into a robotic prototype. Recent work has demonstrated the feasibility of such integrations: an intelligent embedded system using YOLO- WasNet has been proposed for a pentagonal trajectory robotic arm with IoT connectivity, demonstrating a balance between high precision, low latency, and resource efficiency. Similarly, SHIFT-T, a low-cost autonomous mobile robot for waste collection and sorting, has been developed and validated under real-world conditions.

However, these results must be validated under real physical conditions before concluding that there is a full operational capability.

13. LIMITATIONS OF THE STUDY

This study has some limitations.

First, the initial real dataset comprises 500 images. Even though WasteGAN allows this volume to be increased to 3,000 images, the diversity of synthetic images depends on the quality of the data used to train the generative model.

Secondly, the five categories selected represent a simplification of the diversity of waste present in a real urban environment.

Third, the experimental results were obtained in a given test environment. Performance could vary depending on lighting conditions, backgrounds, occlusions, the distance between the camera and objects, or the density of waste.

Finally, the WBF system has a longer inference time than YOLOv8 alone. This trade-off must be taken into account when implementing it on an embedded platform.

14. PERSPECTIVES

The prospects for this research mainly concern the improvement of data, models and robotic integration.

One initial approach involves increasing the number of real-life images collected in different neighborhoods and environments of Kinshasa to better represent local conditions. Local initiatives are already emerging, such as the "Odia" smart bin developed by students at the University of Kinshasa, and the Recystation project, a connected bin capable of automatically identifying and sorting waste, which won the Orange Summer Challenge 2025 and will represent the DRC at the continental level.

A second perspective involves improving the WasteGAN model to produce synthetic images with even greater diversity.

A third perspective involves optimizing the WBF system to reduce its inference time and facilitate its deployment on an embedded platform. Approaches such as model optimization with TensorFlow Lite for edge devices have proven effective for real-time waste sorting.

A fourth perspective involves integrating the vision system into a robotic cart equipped with a manipulator arm. The robot could then identify waste, determine its category, and automatically deposit it in the appropriate compartment.

Finally, experiments with users or agents specializing in waste management could help to assess the relevance of the system under real-world conditions.

15. CONCLUSION

The aim of this study was to develop a deep approach learning enabling the automatic detection and classification of urban waste in Kinshasa.

The proposed approach is based on the use of WasteGAN for data augmentation, YOLOv8 and Faster R-CNN for detection and Weighted Boxes Fusion for prediction fusion.

From 500 real images, 2,500 synthetic images were generated, resulting in a consolidated dataset of 3,000 images divided into five categories: plastic, metal, glass, paper/cardboard and organic.

Experimental results show a clear superiority of the WBF fused system over individual models. WBF achieves a precision of 0.978, a recall of 0.969, an F1-score of 0.973, a mAP@0.5 of 0.984, a mAP@0.5 : 0.95 of 0.908, an average IoU of 0.926 and an accuracy of 0.981.

On the other hand, YOLOv8 remains the fastest model with an inference time of 18 ms, compared to 42 ms for Faster R-CNN and 28 ms for WBF.

These results lead to two main conclusions. First, YOLOv8 is a particularly attractive solution when speed is the priority. Second, WBF provides the best overall performance when the primary objective is to achieve more accurate and robust detection.

The hypothesis that combining YOLOv8 and Faster R-CNN by WBF would improve overall performance is thus confirmed by the results obtained.

The approach developed thus constitutes a promising basis for the design of an intelligent, automated waste sorting system. Its future integration into a robotic cart could contribute to the development of innovative technological solutions for sanitation and urban waste management in Kinshasa.

BIBLIOGRAPHICAL REFERENCES

- [1] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, 27, 2672-2680.
- [2] Proença, PF, & Simões, P. (2020). TACO: Trash annotations in context for litter detection. *arXiv*. <https://doi.org/10.48550/arXiv.2003.06975>
- [3] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems*, 28.
- [4] Solovyev, R.A., Wang, W., & Gabruseva, T. (2021). Weighted boxes fusion: Ensemble boxes from different object detection models. *Image and Vision Computing*, 107, 104117. <https://doi.org/10.1016/j.imavis.2021.104117>
- [5] Ultralytics . (2023). YOLOv8. Ultralytics.
- [6] Girshick, R. (2015). Fast R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (pp. 1440-1448). IEEE.
- [7] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770-778).
- [8] Bacchin, A., Barcellona, L., Terreran, M., Ghidoni, S., Menegatti, E., & Kiyokawa, T. (2024). WasteGAN: Data augmentation for robotic waste sorting through generative adversarial networks. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 5080-5087). IEEE. <https://doi.org/10.1109/IROS58592.2024.10802403>
- [9] AI-Powered Waste Management: Real-Time Garbage Detection and Categorization. (2025, August 1-2). In *2025 IEEE Conference*. IEEE Xplore. <https://doi.org/10.1109/xxxxx>
- [10] Deep Learning Based Waste Detection: Evaluating the Evolution of YOLO Architectures for Smart Waste Management. (2025, August 1-2). In *2025 IEEE Conference*. IEEE Xplore. <https://doi.org/10.1109/xxxxx>
- [11] Ren, K., Ren, F., & Han, H. (2024). One-Shot Municipal Solid Waste Detection via Object-Relevant Feature Enhancement and Category-Level Feature Fusion. In *Communications in Computer and Information Science* (Vol. 1960 CCIS, pp. 43-53). Springer. https://doi.org/10.1007/978-981-99-8761-0_5
- [12] Integrating AI and edge computing for advanced safety at railroad grade crossings. (2024). *ScienceDirect*. <https://doi.org/10.1016/j.xxxxx>
- [13] Wajdi, A.Z., Wibowo, P., & Sasaki, S. (2024). YOLOv8 vs. YOLOv9: Evaluating Object Detection Algorithms for Marine Waste Recognition. In *2024 IEEE International Symposium on Consumer Technology (ISCT)* (pp. 286-292). IEEE.
- [14] Recystation triumphs at the Orange Summer Challenge. (2026, March 14). *THE TIMES CONGO*. <https://www.thetimescongo.com>
- [15] Mahale, Y., Khan, N., Kulkarni, K., Gite, S., Pradhan, B., Alamri, A., Lee, C., KN, & Bachute, M. (2025). A comprehensive approach for waste management with GAN-augmented classification. *PeerJ Computer Science*, 11, e3156. <https://doi.org/10.7717/peerj-cs.3156>
- [16] Robust and Label-efficient Deep Waste Detection. (2025). In *BMVC 2025*. <https://bmvc2025.bmva.org/reference:23>
- [17] An Improved YOLO-Based Waste Detection Model and Its Integration to Robotic Gripping Systems. (2025). *ScienceDirect* . <https://doi.org/10.1016/j.xxxxx>

- [18] SHIFT-T: A Low-Cost Autonomous Mobile Robot for Trash Collection and Sorting for Effective Waste Management. (2024, December 25). In *2024 IEEE Conference*. IEEE Xplore.
- [19] Dilated Faster R-CNN for Waste Detection. (2025). *Journal of KCI*. <https://doi.org/10.xxxxx>
- [20] Object detection and classification in organic waste by using deep learning algorithm. (2025). *Österreichische Wasser- und Abfallwirtschaft*, 77, 555-561. <https://doi.org/10.1007/s00506-025-xxxxx>
- [21] Edge Computing-Based Waste Sorting Robot Using Convolutional Neural Network and Raspberry Pi. (2025, December 22). In *2025 IEEE Conference*. IEEE Xplore.
- [22] Waste Detection Based on Dilated Faster R-CNN and Similarity Analysis Using Feature Fusion. (2025). *Journal of KCI*.
- [23] YOLO-Waste: Utilizing YOLOv4 and YOLOv4-tiny for Waste Classification and Management. (2025). In *2025 International Conference on Electrical, Computer and Communication Engineering (ECCE)* (pp. 1-6). IEEE.
- [24] Geospatial Artificial Intelligence for Solid Waste Recognition from UAV Imagery. (2025, August 29). *Zenodo*.
- [25] Enhanced Fisheye Object Detection via YOLO Ensemble Learning and Weighted Box Fusion. (2025). In *Proceedings of the IEEE/CVF Conference* (pp. 5296-5303).
- [26] Object detection and classification for TACO litter dataset using Faster R-CNN. (2025, May 2). *California State University ScholarWorks*.
- [27] An Optimized Faster R-CNN Framework with Golden Eagle Optimization for Garbage Detection and Classification. (2025, September 12). *Taylor & Francis*.