

Waste Classification using YOLO on Raspberry PI

Vinay Jadhav

Computer Science and Engineering Department
Bharatiya Vidya Bhavan's S.P.I.T.
Mumbai, India

Nathan Cardoso

Computer Science and Engineering Department
Bharatiya Vidya Bhavan's S.P.I.T.
Mumbai, India

Dhruvil Patel

Computer Science and Engineering Department
Bharatiya Vidya Bhavan's S.P.I.T.
Mumbai, India

Abstract— In conjunction with continuous efforts to reduce the consumption and usage of non-recyclable products, efficient waste management requires adequate segregation and recycling of waste. In this paper, we employ computer vision and deep learning techniques to appropriately identify and classify waste among different groups, namely recyclable, non-recyclable, and hazardous waste. We trained three different models of the ultralytics YOLO (You Only Look Once) family, namely, YOLOv5, YOLOv7, and YOLOv8 on trash datasets to identify objects. With the help of YOLOv8, we obtained the highest mAP (Mean Average Precision) value of 83.7% during the training phase. YOLOv5 and YOLOv7 had precision values of 51.4% and 62.43%, respectively. The YOLOv8 model was then executed on a Raspberry Pi microcontroller. We used a web camera as an input device to capture videos of different objects. The microcontroller also controls a 5 V servo motor to separate waste among different bins. This arrangement proves the feasibility for actual segregation systems using YOLO models. This work can be built upon by scaling up the hardware, which could enable real-life deployment.

Keywords— deep learning, computer vision, waste classification, YOLO, Raspberry Pi

I. INTRODUCTION

In today's modern life, effective waste management is very important. With the advancement of technology, there is an increase in the production of goods and agricultural produce, and it positively impacts the lifespan of human beings, increasing the population. This leads to a tremendous increase in waste around us. Traditional waste management often struggles with the proper sorting of waste. This leads to recycling problems and poor resource recovery. The improper management of waste handling, segregation, processing, or decomposition affects environmental health, people's wellbeing, and resource preservation. Thus, there exists a need to provide smarter solutions and better handling of waste. This study presents a new way to segregate waste using modern tools and techniques. These include use of computer vision tools, deep learning techniques, and embedded systems. Computer vision helps digital systems to capture, interpret, and

process visual information from the surrounding world. This capability allows systems to perform complex tasks such as object detection and object classification. Using deep learning, computers learn from data [1]. Transfer learning also plays an important role in such tasks. Transfer learning reuses knowledge from one task to help solve another [2]. This research uses it to refine waste sorting methods. Automated algorithms and micro-controller device help sort materials into recyclable and non-recyclable categories.

- **Recyclable waste:** This includes any kind of waste that we can process and reuse. For example, used or broken glass items can be melted to create recycled glass which can be used to make bottles and other items. Used paper and clothes can be reduced to fibers and new paper can be made. Aluminum and steel cans are also recyclable. As we can see, there are a lot of household items that we can recycle, but we don't send for recycling due to lack of knowledge about recyclable and non-recyclable objects.
- **Non-recyclable waste:** This waste cannot be reused and instead is sent to landfills to be dumped or incinerated. This waste is responsible for the waste management problems we face today. This category of waste cannot be reused and simply must be sent to landfills. Most plastics fall into this category. Today, companies are actively trying to make the most of their packaging recyclable to reduce the amount of non-recyclable waste generated.
- **Hazardous waste:** This category includes materials that pose significant threats to human health and the environment due to their dangerous properties. Hazardous waste is characterized by toxicity, reactivity, ignitability, corrosivity, infectiousness, or radioactivity. Examples include industrial chemicals, medical waste, batteries, electronic waste, pesticides, and certain household products like cleaning solvents and paints. These materials require specialized handling, treatment, and disposal methods to prevent environmental contamination.

II. LITERATURE REVIEW

Addressing waste classification has involved a wide array of algorithms, encompassing everything from conventional machine learning methods like Support Vector Machines (SVM) to the more recent, powerful strides in deep learning. In the following, we summarize various approaches used to solve the waste classification problem over the past decade.

Sakr et al. [3] conducted a comparative analysis between SVM and CNN for three-category waste classification involving paper, plastics, and metal. The SVM demonstrated superior performance with an accuracy of 94.8%, compared to 83% achieved by the CNN. The SVM implementation was carried out on a Raspberry Pi 3.

Chu, Huang et al. [4] proposed a Multilayer Hybrid System (MHS) for waste classification, integrating CNN-based feature extraction, feature engineering and an MLP classifier, achieving 92% accuracy.

Adedeji and Wang [5] utilized the ResNet-50 architecture for feature extraction, the support vector machine for classification. On the Trashnet dataset, their approach showed an accuracy of 87%.

Aral, Keskin et al. [6] trained multiple deep learning models on the Trashnet data set and compared their accuracies. The fine-tuned DenseNet121 and InceptionResNetV2 models demonstrated accuracy up to 95%, highlighting the efficacy of deep learning.

Vo, Son et al. [7] utilized transfer learning to develop the novel DNN-TC architecture, which was trained on both the TrashNet and VN-Trash, resulting in accuracies of 94% and 98%, respectively.

Bobulski and Kubanek [8] developed a 15-layer CNN capable of classifying plastic waste into different categories. They achieved an accuracy of 98% using images sized 120 by 120 pixels, which significantly reduced learning time.

Mao et al. [9] trained a modified version of the DenseNet121 on the Trashnet dataset, utilizing a genetic algorithm to optimize the fully connected layers. This approach achieved an accuracy of 99.6%.

Thokrairak et al. [10] trained the MobileNet SSD model to classify plastic bottles, metal cans, and glass bottles with accuracies of 95%, 86%, and 82%, respectively.

Chiu et al. [11] proposed the MobileNet-SSD v2 model, which achieved a mean average precision (mAP) of 75.9 on the VOC data set.

Ma et al. [12] utilized the MobileNetv2 model for waste classification, achieving an accuracy of 98.7% with a processing speed of approximately 70 ms.

Zhang et al. [13] used transfer learning to retrain a DenseNet169 model on their own dataset, which produced an accuracy of 82.8% for the classification of five categories of waste.

Mazir Mohammed et al. [14] implemented a neural network utilizing four different feature extractors, training a separate neural network for each feature. Final classification was performed based on majority voting among the four networks, resulting in a three-way classification between paper, metal, and other trash with an accuracy of 91.7%.

Sirawattananon et al. [15] introduced an IoT-based waste sorting system using the ResNet-50 model, achieving 98.81% accuracy. The model, trained on more than 5,000 images, was

deployed on a Raspberry Pi 3 Model B with a servo motor for segregation. The system used a conveyor belt and a 20 kg torque servo motor to segregate waste into different bins.

Susanth et al. [16] performed a comparative study of ResNet50, DenseNet169, and VGG16 on a dataset containing more than 4,100 images, observing better performance with the DenseNet169 model.

Bawankule et al. [17] trained various YOLO models on the dataset titled 'Classification model for waste materials in residential areas' [18], achieving an mAP of 95.4% with YOLOv7 and 97.7% with the YOLOv8 model. The dataset used contained 9,800 images.

Tian, Shi et al. [20] developed an improved MobileNetV3 model, which uses the CBAM module for feature enhancement and Mish activation. The model achieved a precision of 96.55% and reduced the number of parameters by 56.6%.

Although significant advances have been made in computer vision for object detection and classification, limited research has focused on integrating these models with hardware systems. This gap is the focus of our research.

III. METHODOLOGY

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

The classification of waste according to its recyclability is crucial for effective waste segregation, allowing a higher proportion of generated municipal solid waste to be recycled. Our goal is to leverage computer vision and deep learning techniques to accurately classify common waste items into three categories: recyclable, non-recyclable, and hazardous. The chosen deep learning model will be deployed on a Raspberry Pi microcontroller to ensure efficient and real-time waste classification.

To achieve this objective, we have selected the YOLO (You Only Look Once) family of object detection models, which utilize the CSPDarknet as a backbone. YOLO models are specifically designed to perform both object localization and detection in a single shot, offering a significant advantage over other architectures such as VGG-16 and MobileNet. Although VGG-16 is computationally intensive and unsuitable for hardware implementations, MobileNet, although lightweight and efficient for classification, requires additional SSD layers for proper detection. In contrast, the YOLO model is optimized for speed and accuracy, which makes it suitable for real-time applications on devices such as the Raspberry Pi.

The architecture of YOLO models is organized into three primary components: the backbone, the neck, and the head. Each component serves a distinct purpose within the overall framework, although the algorithms and methodologies employed may vary between different YOLO versions.

- **Backbone:** The backbone extracts features from the raw input image, forming the foundation of the model's feature representation
- **Neck:** The neck aggregates features from multiple scales, facilitating robust multiscale detection. This component improves the ability of the model to detect objects of varying sizes.
- **Head:** The head is responsible for generating predictions, including bounding box coordinates and confidence scores for detected objects.

Using the YOLO model architecture, our aim is to balance computational efficiency and high accuracy, ensuring that our waste classification system can be seamlessly integrated with hardware for real-time applications.

A. Architecture of YOLOv5

YOLOv5 uses CSPDarknet53 as the backbone. CSPDarknet53 combines the Darknet network with cross-stage partial connections (CSP). CSP divides the base feature net into two parts. One part would pass through the dense layers, and the other part is directly concatenated. This helps maintain the richness of features extracted while improving computational efficiency. Spatial Pyramid Pooling helps the model capture multiscale features. Spatial Pyramid Pooling (SPP) increases speed by combining many different features into a feature map.

The neck of the model is used to combine or aggregate many features. PANet combines features of many different levels, which allows for multiscale detection. The two main techniques involved in PANet are bottom-up path augmentation and adaptive pooling. They help the model prioritise the best features and perform multi-scale detection. PANet is a major leap in feature fusion compared to traditional feature pyramid networks. The structure of PANet is seen in Fig. 1. It is still very computationally efficient.

The model head is ultimately responsible for performing the actual object detection. It takes advantage of the rich features provided by PANet and uses three different convolutional layers to generate final predictions at different scales. The image is split into grids, and predictions are made for each of the grids in the image. YOLOv5 uses dynamic anchor boxes

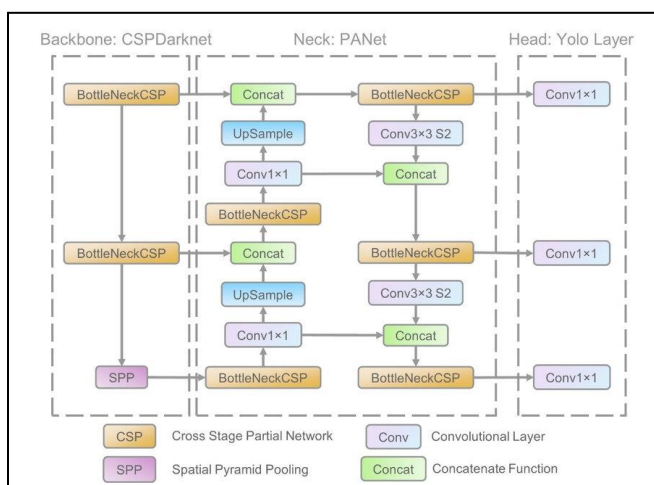


Fig 1: Architecture of the YOLOv5 model [21]

B. Architecture of YOLOv7

The YOLOv7 model is a significant improvement from the YOLOv5 model discussed previously. As shown in Fig. 2, the backbone of the network uses some of the concepts introduced by Darknet, like residual connections and convolutional layers, but the architecture is very different. The backbone is based on ELAN (Efficient Layer Aggregation Network) which was introduced in YOLOv6. ELAN also performs splitting and aggregation like the CSPDarknet but has multiple layer aggregation techniques. ELAN was introduced in YOLOv6. The YOLOv7 backbone uses E-ELAN (Extended ELAN) which is an improved version of ELAN. It further refines the aggregation techniques and also integrated attention mechanisms. E-ELAN provides a faster inference than ELAN. The neck of YOLOv7 also uses a PANet, just like YOLOv5. However, YOLOv7 uses a much more refined and advanced version of the PANet network that has better feature fusion and dynamic aggregation. In addition, the use of E-ELAN makes it even more efficient. Hence there is a better performance, especially in the case of smaller objects. The prediction head in YOLOv7 uses an anchor-free approach to determine the bounding boxes. It makes the model more flexible, allowing it to handle a wider range of objects for detection. The improved loss function provides a better balance between accuracy and speed.

C. Architecture of YOLOv8

The YOLOv8 backbone sees the return of the CSPDarknet architecture, as both YOLOv6 and YOLOv7 did not use this and preferred an ELAN-based backbone. The CSPDarknet used in YOLOv8 has been optimized with adjustments and changes in the number of layers and channels. This means that YOLOv8 is much faster at extracting rich features from images.

As shown in Fig. 3, the neck of the YOLOv8 architecture uses PANet for bottom-up feature aggregation. The newer PANet used in YOLOv8 performs enhanced aggregation, improved feature propagation, and more effective multiscale fusion

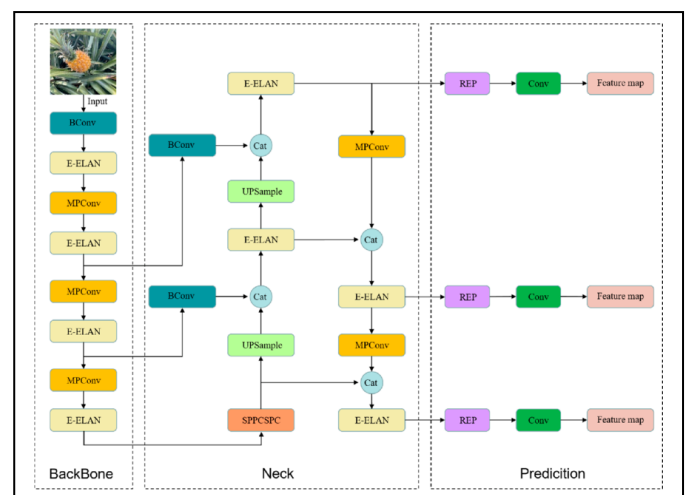


Fig 2: Architecture of the YOLOv7 model [21]

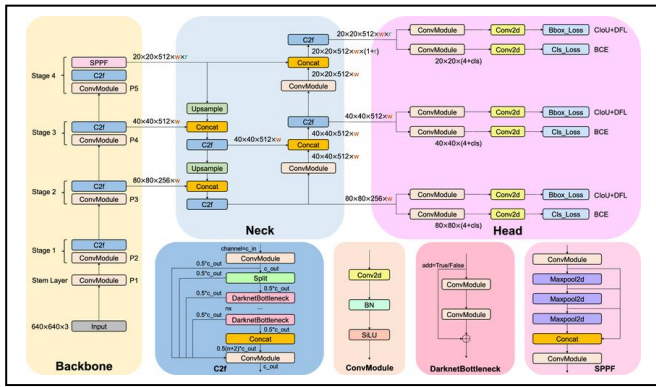


Fig 3: Architecture of the TOLov8 model [22]

Both the backbone and neck of the YOLOv8 make use of C2f blocks instead of the regular CSPDarknet bottlenecks.

C2f stands for crossover layers with second-order fusion. Crossover layers perform the integration of feature maps at different levels of the network. Second-order fusion refers to the combination of spatial and channel-wise information. C2f blocks make multi-scale detection much easier, while also resulting is faster training and inference.

YOLOv8 does not use an anchor bounding box like YOLOv5. It can detect without any anchor boxes, enhancing its adaptability to various objects with different shapes and sizes. This model has enhanced performance and accuracy. YOLOv8 also has adaptive loss functions that aid in better training. The three models will be trained and tested to determine which one gives the best results. The best-performing model is executed on the Raspberry Pi 3 for performing segregation.

Fig. 4 showcases a block diagram of the proposed segregation system. The Raspberry Pi was chosen due to its superior processing power, which makes it apt for real-time inference. It also uses a Debian-based operating system, which is ideal for our project as packages like OpenCV and PyTorch are essential. Thus, Raspberry Pi was deemed to be the best choice for deployment of YOLO models.

For camera interfacing, a USB webcam will be connected to the Raspberry Pi. As the servomotors utilised are quite small

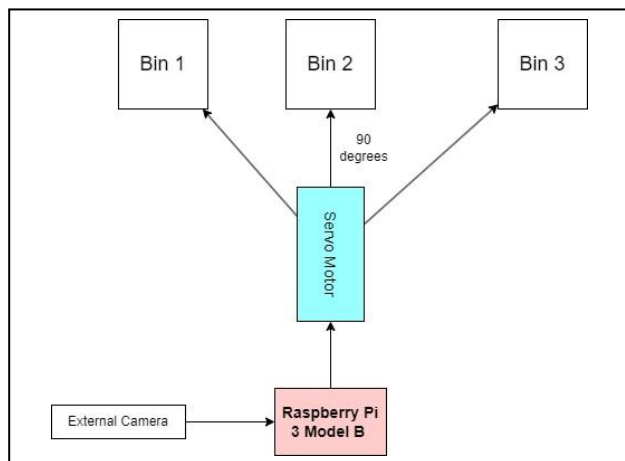


Fig 4: Schematic of the proposed system

and can handle a limited weight, the prototype is meant to provide feasibility proof rather than a product ready for deployment into the real world.

IV. EXPERIMENTATION AND RESULTS

We have utilized the concept of transfer learning, which involves fine-tuning a pretrained model on a custom dataset to improve its performance. Traditionally, machine learning models require that the data and features used for training reside in the same feature space. This means that the features extracted from the data are directly used as input to the learning algorithm. However, transfer learning transcends this limitation by operating across different feature spaces. For example, YOLO models are initially pre-trained on the COCO dataset and subsequently fine-tuned using a custom dataset tailored to specific tasks.

The YOLOv5 model was trained on Trashnet [24], TACO [25] and COCO datasets. The YOLOv7 model was trained on the TACO dataset. The YOLOv8 model was trained on a public dataset of about 6700 images [26]. The training was performed in a Google Colab environment using the provided T4 GPU.

As seen in TABLE I, the YOLOv8 model easily outperforms both the YOLOv5 and YOLOv7 models. Hence, this model will be implemented on the Raspberry Pi.





The categorization of the dataset used on YOLOv8 for recyclable, non-recyclable, and hazardous waste is as follows: cardboard boxes, plastic bottles, plastic bottle caps, and reusable paper are all recyclable. Non-recyclable objects are plastic bags, scrap paper, sticks, plastic cups, snack bags, plastic boxes, straws, plastic lids, scrap plastics, cardboard bowls, and plastic cutlery. Hazardous items are chemical spray cans, chemical plastic bottles, chemical plastic gallons, paint buckets, and light bulbs. The accuracy of the models seen in TABLE 1 is pertaining to the recognition of these objects.

For execution of the model on a Raspberry Pi, we created a separate Python virtual environment. Afterwards, we installed all the required packages and their dependencies in that environment. A Python script is written to capture an image every 20 seconds, detect the object using YOLOv8, and then rotate the motor accordingly. It takes approximately 20 seconds to detect and classify the image. In this way, the segregation process is completely automated. The experimental setup in Fig. 5 shows the placement of the web camera, Raspberry Pi 3 and the prototype. The list of components is seen in TABLE II. We tested for three-way classification on the prototype was done using the following objects: Paint bucket, cardboard box, chemical plastic gallon, spray can, light bulb. plastic bottle, plastic bottle cap, plastic spoon, plastic bag and scrap paper were used. The model was able to detect all of these objects accurately under good lighting conditions.

TABLE I. PERFORMANCE COMPARISON OF YOLO MODELS

Model	Dataset Size	Evaluation Metric
YOLOv5	8500	Accuracy = 62.43%
YOLOv7	1500	Accuracy = 51.4%
YOLOv8	6684	mAP = 83.7%

TABLE II. REQUIRED COMPONENTS

Component Name	Photo
Raspberry Pi 3	
Servo Motor	
USB Web camera	
Micro USB cable	

V. EXPERIMENTATION AND RESULTS

In this work, we have successfully performed waste classification using the YOLOv8 deep learning model. The hardware implementation demonstrates the feasibility of using YOLO with Raspberry Pi for waste segregation.

Data inadequacy is a limitation because data sets would not be able to take into account every single object that gets put in the trash. Other challenges include occasional misclassification between plastic and glass due to visual similarity, and imbalances in datasets.

This research can be built on improved hardware. The use of powerful processors like NVIDIA Jetson Nano, which can enable the training of the model based on images captured with a webcam. This would improve both speed and accuracy. The use of either servo motors with higher torque values or DC motors, along with a more robust prototype material, can be explored to build a deployment ready product. Integration of IoT technologies can enable execution on computers, with the results of the classification being sent to a robust hardware system.



Fig 5: Experimental setup

REFERENCES

- [1] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep learning for computer vision: a brief review," Computational Intelligence and Neuroscience, vol. 2018, pp. 1–13, Feb. 2018.
- [2] A. Hosna, E. Merry, J. Gyalmo, Z. Alom, Z. Aung, and Mz. A. Azim, "Transfer learning: a friendly introduction," Journal of Big Data, vol. 9, no. 1, Oct. 2022.
- [3] G. E. Sakr, M. Mokbel, A. Darwich, M. N. Khneisser, and A. Hadi, "Comparing deep learning and support vector machines for autonomous waste sorting," in 2016 IEEE International Multidisciplinary Conference on Engineering Technology (IMCET), Beirut, Lebanon, 2016, pp. 207212.
- [4] Y. Chu, C. Huang, X. Xie, B. Tan, S. Kamal, and X. Xiong, "Multilayer Hybrid Deep-Learning method for waste classification and recycling," Computational Intelligence and Neuroscience, vol. 2018, Article ID 8357502.
- [5] O. Adedeji and Z. Wang, "Intelligent waste classification system using deep learning convolutional neural network," Procedia Manufacturing, vol. 35, pp. 607-612, 2019.
- [6] R. A. Aral, S. R. Keski, M. Kaya, and M. Hacıomeroğlu, "Classification of TrashNet dataset Based on deep learning models," in 2018 IEEE International Conference on Big Data, 2018.
- [7] A. H. Vo, M. T. Vo, and T. Le, "A novel framework for trash classification using deep transfer learning," IEEE Access, vol. 7, no. 3, pp. 178631-178639, December 2019.
- [8] J. Bobulski and M. Kubanek, "Waste classification system Using image processing and convolutional neural networks," in Advances in Computational Intelligence, I. Rojas, G. Joya, and A. Catala, Eds. IWANN 2019.
- [9] W.-L. Mao, W.-C. Chen, C.-T. Wang, and Y.-H. Lin, "Recycling waste classification using optimized convolutional neural network," Resources, Conservation and Recycling, vol. 164, January 2021.
- [10] S. Thokairak, K. Thibuy and P. Jitngernmadan, "Valuable Waste Classification Modeling based on SSD-MobileNet," 2020 - 5th International Conference on Information Technology (IncIT), Chonburi, Thailand, 2020
- [11] Y.C. Chiu, C.Y. Tsai, M.D. Ruan, G.Y. Shen, and T.T. Lee, "MobilenetSSDv2: an improved object detection model for embedded systems," in 2020 International Conference on System Science and Engineering (ICSSE), 2020, pp. 1-5.
- [12] H. Ma, Y. Ye, J. Dong, and Y. Bo, "An Intelligent Garbage Classification System Using a Lightweight Network MobileNetV2," in 7th International Conference on Signal and Image Processing (ICSIP), 2022, pp. 531-535.
- [13] Q. Zhang, Q. Yang, X. Zhang, Q. Bao, J. Su, X. Liu, "Waste image classification based on transfer learning and convolutional neural network," Waste Management, vol. 135, pp. 150-157, 2021.
- [14] M. A. Mohammed, M. J. Abdulhasan, N. M. Kumar, S. Chopra, et al., "Automated waste-sorting and recycling classification using artificial neural network and features fusion: a digital-enabled circular economy vision for smart cities," Multimedia Tools and Applications, vol. 82, pp. 39617-39632, 2023.
- [15] C. Sirawattananon, N. Muangnak, and W. Pukdee, "Designing of IoTBased Smart Waste Sorting System with Image-based Deep Learning Applications," in 8th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2021, pp. 383-387.
- [16] G. S. Susanth, L. J. Livingston, and L. A. Livingston, "Garbage Waste Segregation Using Deep Learning Techniques," IOP Conference Series: Materials Science and Engineering, vol. 1012, p. 12040.
- [17] R. Bawankule, V. Gaikwad, I. Kulkarni, S. Kulkarni, A. Jadhav, and N. Ranjan, "Visual Detection of Waste using YOLOv8," in International Conference on Sustainable Computing and Smart Systems (ICSCSS), 2023, pp. 869-873.
- [18] <https://universe.roboflow.com/thesis-project-sacr3/classification-model-for-waste-materials-in-residential-areas/dataset/3>
- [19] J. Li et al., "Automatic Detection and Classification System of Domestic Waste via Multimodal Cascaded Convolutional Neural Network," in IEEE Transactions on Industrial Informatics, vol. 18, no. 1, pp. 163-173, Jan. 2022

- [20] X. Tian, L. Shi, Y. Luo, and X. Zhang, "Garbage Classification Algorithm Based on Improved MobileNetV3," IEEE Access, vol. 12, pp. 44799-44807, 2024.
- [21] Real-time Face Mask Detection in Video Data - Scientific Figure on ResearchGate. Available from: <https://www.researchgate.net/figure/YOLO-v5-Architecture-Overview-3-fig2-351355008>
- [22] A Pineapple Target Detection Method in a Field Environment Based on Improved YOLOv7 - Scientific Figure on ResearchGate. Available from: <https://www.researchgate.net/figure/The-network-structure-of-the-original-YOLOv7-fig4-368669692>
- [23] AIRCRAFT DETECTION WITH DEEP NEURAL NETWORKS AND CONTOUR-BASED METHODS - Scientific Figure on ResearchGate.<https://www.researchgate.net/figure/YOLOv8-architecture-16-fig3-387434154> [accessed 24 May 2025]
- [24] <https://github.com/garythung/trashnet>
- [25] <http://tacodataset.org/>
- [26] <https://universe.roboflow.com/ai-project-i3wje/waste-detectionvqkjo/dataset/9>