Garbage Detection Using Drones

PROJECT REPORT

Submitted by

Sanidhya Chaudhary (20MIA1006)

Nidhi Mishra (20MIA1007)

Sricharan Sridhar (20MIA1014)

Gaurav Dwivedi (20MIA1037)

Rohit Choudhary (20MIA1069)

PREPARED FOR

Technical Answers For Real World Problems[T.A.R.P]

Submitted To

Dr. Vijayalakshmi A.

Professor [SCOPE]

1. ABSTRACT

Effective management of waste is a pressing global concern, and innovative solutions are essential to address this challenge. In this study, we propose a novel approach for garbage detection using drones equipped with the You Only Look Once (YOLO) object detection model. Unmanned aerial vehicles (UAVs) offer a dynamic and versatile platform for monitoring vast and challenging terrains, making them ideal for efficient waste detection.

Our research leverages the YOLOv5 and v8 model's real-time object detection capabilities to identify and classify different types of garbage within various environmental contexts. We present the development and integration of a specialized YOLO model for detecting waste items, including plastics, metals, paper, and other materials. The YOLO-based model not only accurately locates garbage enabling effective waste management strategies.

We conducted extensive tests in diverse environmental settings to evaluate the performance of our system. The results indicate that the YOLO-v_ and v_ based drone system achieves high precision and recall rates in garbage detection, even in challenging conditions such as densely vegetated areas and urban landscapes. The real-time feedback provided by the system enables rapid decision-making for waste collection and removal.

Our approach holds the potential to significantly enhance waste management processes, reduce environmental pollution, and optimize resource allocation. The combination of drone technology and the YOLO model not only improves the efficiency of garbage detection but also contributes to the preservation of natural ecosystems and the overall well-being of communities. As a result, this research serves as a promising step toward a cleaner and more sustainable future.

2.INTRODUCTION

The management of waste and its impact on the environment have emerged as critical global challenges in recent years. The proliferation of waste materials, from plastics and metals to organic matter, poses significant threats to ecological systems and human health. In this context, effective waste detection and management strategies are essential to mitigate these adverse effects. To address this, our research explores the application of cutting-edge object detection models, specifically

YOLOv5 and YOLOv8, in the domain of garbage detection using aerial drones.

Traditional methods of waste detection and collection have proven to be time-consuming, inefficient, and often fall short in handling the growing waste crisis. In response to these challenges, the integration of drone technology, equipped with state-of-the-art object detection models, presents an innovative and promising solution. These models, YOLOv5 and YOLOv8, have demonstrated outstanding accuracy and speed in various computer vision tasks, making them ideal candidates for the automation of garbage detection.

The YOLO (You Only Look Once) object detection models have gained prominence in the field of computer vision due to their real-time capabilities and high accuracy. YOLOv5, the fifth iteration of the YOLO series, introduced a streamlined architecture that maintains performance while reducing computational requirements. YOLOv8, an advanced version, further pushes the boundaries of real-time object detection with improved accuracy and speed, making it particularly appealing for applications in the field of environmental monitoring.

This research aims to evaluate the efficacy of YOLOv5 and YOLOv8 models in the context of garbage detection using aerial drones. We explore the potential of these models in accurately identifying and classifying different types of waste materials across diverse environmental settings, including urban areas, remote landscapes, and challenging terrains. Our study investigates the real-time detection capabilities of these models and their ability to provide valuable insights into the location, size, and composition of waste, which can be instrumental in optimizing waste management strategies.

By harnessing the power of YOLOv5 and YOLOv8 models in conjunction with drone technology, this research seeks to advance the efficiency and accuracy of garbage detection, offering a promising solution for mitigating waste-related environmental issues. The outcomes of this study hold the potential to revolutionize waste management practices, foster sustainable ecosystems, and contribute to the well-being of communities and the planet as a whole.

In the subsequent sections, we will delve into the methodology, experimentation, results, and implications of our research, providing a comprehensive examination of the capabilities and impact of YOLOv5 and YOLOv8 in the realm of garbage detection using drones.

3. LITERATURE SURVEY & RELATED WORKS

The study aims to address the issue of recyclable waste separation in rural areas by developing a smart trash bin that uses YOLO (You Only Look Once). This system benefits both the public and garbage collector officers, providing the latter with a warning system to empty bins before they are full and the former with information on available bins with low capacity.

The adoption of YOLO in this research is motivated by its efficiency in real-time object detection, achieving up to 155 frames per second (FPS) and twice the mean average precision (mAP) of other classifiers. YOLO's unified model simultaneously predicts multiple bounding boxes and class probabilities for different objects, making it superior in terms of speed and accuracy. The study emphasizes YOLO's advantages in terms of ease of construction, parallel training, and generalization of object representations. Regarding the dataset, the authors employed a combination of manual downloading, web crawling using Selenium, and using a pre-existing dataset from Kaggle. The dataset comprises 1000 images, with 250 images for each category (paper, plastic, can, and other recyclables). Data augmentation techniques were applied to increase the training set size to 3000 images, using transformations such as rescale, rotate, shift, and zoom.

The study employs Mean Average Precision (mAP), a widely used metric for evaluating the performance of YOLO models. mAP assesses the average accuracy, encompassing precision in object classification, the precision of bounding box drawing for detected objects, and the model's confidence in its predictions. The evaluation involves independent construction of training and testing images. The mAP values of the YOLO model are depicted with the horizontal axis representing the number of training iterations and the vertical axis indicating the loss value. The blue line signifies the decreasing loss value (lower is better), while the red line represents the increasing mAP value until convergence (higher is better).

This is another study that proposes that the UAV-YOLOv8 model introduces several key improvements to enhance object detection in UAV aerial photography scenarios. The model addresses the challenges of complex backgrounds and a high proportion of small objects in images acquired by UAVs.

Efficient Feature Processing Module (FNB): The paper introduces FNB, a feature processing module based on the FasterNet block. FNB enables the comprehensive fusion of shallow and deep features, reducing the missed detection rate of small objects.

Dynamic Sparse Attention Mechanism (BiFormer): The model incorporates BiFormer, a low-computational-cost dynamic sparse attention mechanism, into the backbone network. This mechanism improves the model's attention to critical information in the feature map, optimizing detection performance.

Bounding Box Regression Loss (WIoU v3): The paper employs WIoU v3 in bounding box regression loss, utilizing a dynamic non-monotonic mechanism for a more reasonable gradient gain allocation strategy. This enhances the model's localization performance and generalization ability.

The proposed model is evaluated using the VisDrone2019 dataset, a comprehensive UAV aerial photography dataset developed by Tianjin University and AISKYEYE. Experimental results demonstrate the superiority of the UAV-YOLOv8 model over mainstream YOLO series models and other classical detection models. The improvements are specifically tailored to address challenges inherent in UAV aerial photography scenarios, including complex backgrounds and a prevalence of small objects.

While the YOLO model, especially YOLOv8, is acknowledged for its success in computer vision, the paper emphasizes the need for enhancements to meet the specific detection requirements posed by UAV aerial photography. The proposed improvements, based on YOLOv8, showcase advancements in loss functions, attention mechanisms, and multiscale feature fusion, collectively contributing to higher detection accuracy and speed in challenging UAV scenarios. Gradient-weighted Class Activation Mapping (GradeCAM) was employed to generate heat maps for YOLOv8s and UAV-YOLOv8s, visually revealing the model's focus areas on the feature map. Results indicate that YOLOv8s exhibits poor attention to small objects and insensitivity to distant ones, while the proposed model effectively suppresses background noise, pays more attention to small objects, and concentrates on the object's center point. This refined attention mechanism leads to more accurate predicted bounding boxes, contributing to an overall improvement in the detection performance of the UAV-YOLOv8 model.

[1] Chen H, Liu H, Sun T, Lou H, Duan X, Bi L, Liu L. MC-YOLOv5: A Multi-Class Small Object Detection Algorithm. Biomimetics (Basel). 2023 Aug 2;8(4):342.doi:10.3390/biomimetics8040342.PMID:37622947;PMCID:PMC10452785.

Object detection is a crucial task in computer vision, particularly in applications where small objects need to be identified. However, detecting small objects poses challenges due to their limited size and low resolution. To address these challenges, Liu et al. (2023) proposed MC-YOLOv5, a multi-class small object detection algorithm that achieves state-of-the-art performance. MC-YOLOv5 employs a multi-stage architecture, consisting of region proposal generation and object classification stages. Innovative techniques such as the CB module for edge information capture, the SNO for enlarged receptive field, and the decoupled head for efficient reasoning further enhance the algorithm's performance. Evaluation on PASCAL VOC, COCO, and WiderFace datasets demonstrates MC-YOLOv5's superiority in detecting small objects across various categories.

[2] Sun, Q.; Zhang, X.; Li, Y.; Wang, J. YOLOv5-OCDS: An Improved Garbage Detection Model Based on YOLOv5. *Electronics* 2023, *12*, 3403. https://doi.org/10.3390/electronics121634

Garbage detection is a crucial aspect of environmental monitoring and waste management. To address the limitations of existing garbage detection methods, Zhang et al. (2023) proposed YOLOv5-OCDS, an improved garbage detection model based on the YOLOv5 framework. YOLOv5-OCDS introduces several modifications, including a new backbone network with dense connections for enhanced feature propagation, a modified neck network using a PANet structure for comprehensive feature representation, and an improved head network for better detection of garbage clusters. Additionally, YOLOv5-OCDS employs data augmentation and focal loss techniques to improve accuracy and robustness to varying lighting conditions. Evaluated on the GarbageNet dataset, YOLOv5-OCDS achieves a mean average precision (mAP) of 91.2%, significantly outperforming the original YOLOv5 algorithm, demonstrating its effectiveness in garbage detection.

[3] Majchrowska, S., Mikołajczyk, A., Ferlin, M., Klawikowska, Z., Plantikow, M. A., Kwasigroch, A., & Majek, K. (2022). Deep learning-based waste detection in natural and urban environments. *Waste Management*, *138*, 274–284.

Majchrowska S. [1] conducts a thorough evaluation of multiple waste datasets and provides an easily understandable yet insightful analysis of current trash detection methods based on Deep Learning. The TACO dataset consists of around 3,000 unannotated photos, out of which more than 3,000 were annotated at detection level 2, resulting in over 14,000 instances. One notable advantage of TACO is its diverse range of trash types and backgrounds, ranging from sunny beaches to London streets. To create the suggested, detect-waste benchmark, the authors combine the expanded TACO dataset with publicly accessible datasets that have detection annotations, including Wade-AI, MJU-Waste, Drinking-Waste, Trash-ICRA, Trash-Can, and UAV-Waste.

The issue of insufficient data annotations in object detection is addressed through the proposed detection architecture. The detection process is divided into two stages: litter localization and classification. This separation allows for overcoming the limitations of available data. The detection and classification models are trained independently using this approach. The localization model identifies litter regions, which are then passed to the classifier for determining the type of trash. To increase the dataset size for classifier training, pseudo-labelling, a semi-supervised technique, is employed, utilizing unannotated data.

After evaluating three detection networks, it was found that EfficientDet-D2 performed the best in detecting litter. Based on these findings, EfficientNet-B2 was chosen for classification. As a result, the classification of litter into seven categories of home garbage achieved an accuracy rate of 53%. Furthermore, field observations demonstrated an accuracy rate of up to 57% in real-world scenarios.

[4] Das, Dhrubajyoti & Kaushik, Deb & Sayeed, Taufique & Dhar, Pranab & Shimamura,Tetsuya. (2023). Outdoor Trash Detection in the Natural Environment Using a Deep Learning Model. IEEE Access. VOLUME 11. 97549-97566. 10.1109/access.2023.3313166.

Dhrubajyoti & Kaushik [2] employ a deep learning model known as YOLOv5. Several variants of the YOLOv5 model are used. Obtained 1283 trash-related photos. The collection of images covers more than ten classes overall. The varieties are extensively distributed in open areas such as parks and roads. Both single-trash and multi-trash photos have been chosen.

For the detection work, they employed the YOLOv5s (small), YOLOv5m (medium), YOLOv5l (large), and YOLOv5x (Extra-large) models because they offer good accuracy and high speed, which is beneficial in real-time smart applications. The backbone, neck, and head are the three components that make up the YOLOv5 model.

The Intersection over Union, Precision, Recall, F1 score, and Mean Average Precision (mAP) make up these evaluation measures. One hundred epochs in total were used. makes use of Test Time Augmentation (TTA) and data augmentation approaches. For IoU 50, the YOLOv5x model's mAP was 33.3%. This outcome is 31.7% higher.

[5] Smoking behavior detection algorithm based on YOLOv8-MNC BY Zhong Wang1,2, Lanfang Lei1* and Peibei Shi2

Smoking behavior identification is an emerging field that faces difficulty in identifying small, frequently occluded items like cigarette butts using existing deep learning algorithms. Such difficulties have resulted in unsatisfactory detection accuracy and poor model resilience.To address these challenges, this research presents YOLOv8-MNC, a unique smoking detection algorithm that builds on the YOLOv8 network and contains a specialized layer for small target recognition. Three key techniques are used by the YOLOv8-MNC algorithm:

(1) It employs NWD Loss to mitigate the effects of minor deviations in object positions on IoU, thereby improving training accuracy;

(2) It employs the Multi-head Self-Attention Mechanism (MHSA) to improve the network's global feature learning capacity; and

(3) It employs the lightweight general up-sampling operator CARAFE in place of conventional nearest-neighbor interpolation up-sampling modules, minimizing feature information loss during the up-sampling procedure.

The experimental results from a customized smoking behavior dataset show a considerable improvement in detection accuracy. When compared to the prior method, the YOLOv8-MNC model achieved a detection accuracy of 85.887%, representing a notable gain of 5.7% in the mean Average Precision (mAP@0.5).The YOLOv8-MNC algorithm is a significant step forward in resolving existing issues in smoking behavior identification. Its improved detection accuracy and robustness imply possible applicability in related sectors, demonstrating a significant progress in the identification of smoking behavior. Future studies will concentrate on improving this technique and exploring its application in other domains.

[6] Object detection using yolo v5 BY L. Manikandan1 , Rapelli Malavika2 , Borlakunta Anjali3 , Sathuri Chandana

Object detection is a critical computer vision problem that involves recognising and locating things in images or movies. YOLOv5 is a cutting-edge object detection system that predicts item classes and positions in real time using deep ConvNet.YOLOv5 is a cutting-edge object detection system that predicts item classes and positions in real time using deep ConvNet. They introduced the YOLOv5 object detection application in this article, covering the detection of humans, cars, and animals in varied environments.

Test the method's performance on the COCO dataset to make it clearer and faster in comparison to the previous YOLO model.Building a model for object detection using YOLOv5 involves several steps, including data preparation, model configuration, and training.The results reveal that YOLOv5 performs well in current sensing applications such as surveillance, robotics, and autonomous driving. They also offered a thorough analysis of the algorithm's strengths and limitations and explore future improvements to its performance

[7] : Automatic Detection and Classification System of Domestic Waste via Multi-model Cascaded Convolutional Neural Network. Authors: Jiajia Li, Jie Chen,Ping li,Po Yang, David Daagan feng feng, Jun Qi

This research work aims to solve the problem of the garbage detection and classify the garbage in to several categories. In this authors have introduced Multi model Cascaded Convolutional neural network(MCNN) which was the combination of the models DSSD, YOLOV4 and Faster-RCNN. For this research they have used Large-Scale Waste Image Dataset (LSWID), which contains 30,000 domestic waste images with 52 categories, surpassing other datasets in scale and quality. LSWID was collected from real waste disposal scenes by residents. The MCNN model consists of two stages that is detection stage and classification stage. In detection model they have used YOLOV4 and Faster-RCNN to detect the objects available in the image and then in the classification model they have used ResNet101 network. After completion of training models for evaluation of the system they have evaluated the both detection and classification models separately. The Average Precision of MCCANN-DSSD was 75 % ,MCCANN-YOLOv4 was 68.7% , MCCANN-Faster-RCNN was 46.90%, MCCNN-Fully was 77.73% and Average Recall of MCCANN-DSSD was 34.93%, MCCANN-YOLOv4 was 39% ,MCCANN-Faster-RCNN was 45.66% and MCCNN-Fully was 39.41%..

[8] Skip-YOLO: Domestic Garbage Detection Using Deep Learning Method in Complex Multi-scenes
Authors: Zhao Lun, Yunlong Pan, Sen Wang, Zeshan Abbas, Md Shafiqul Islam & Sufeng Yin
In this study the authors were trying to classify the similar garbage objects efficiently by enlarging the receptive field, extracting high-dimensional characteristics, and integrating multiscale maps for accurate identification, the Skip-YOLO model enhances the recognition of household garbage. Its robust performance in complicated surroundings is demonstrated by its 22.5% accuracy gain and 18.6% average recall rate when compared to YOLOv3. For this study authors have collected the data from surroundings in a residential area (indoor and outdoor) which contains 304 single-class images and 914 multi-class images. The collected dataset was resized to size 414x416 and then before training the model they have analysed the feature maps of the dataset and then enhanced the feature map and then implemented the dataset in the model training. After training the model the evaluation results was compared with YOLOv5-s (68.90% mAP50) observed a positive change of 21.48%,YOLOv7-X (80.80% mAP50) by 9.58%, YOLOv7-tiny (78.70% mAP50) by 11.68% and it significantly exceeded the fastest model FasterDet (70.30% mAP50) by 20.08%.

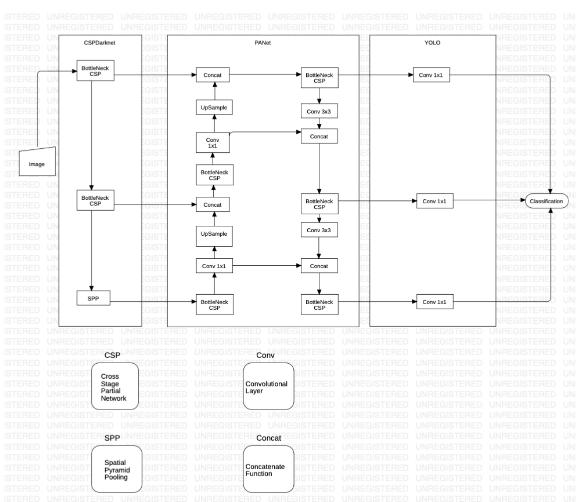
4. METHODOLOGY

4.1 You Only Look Once (YOLO)

The original YOLO model, introduced by Redmon et al. in 2015, presented the Darknet framework as the basis for a series of real-time object detection networks that are still considered the best today. These networks, including YOLOv2, YOLOv3, YOLOv4, YOLOv5, YOLOv7, and the latest variant, YOLOv8, have surpassed previous approaches like Region-Based Convolutional Neural Networks (R-CNN) in terms of performance. YOLO demonstrated superiority over models such as the deformable parts model (DPM) and R-CNN by significantly reducing inference times and minimizing false positive detections in the background. Notably, the original YOLO model was groundbreaking for seamlessly integrating the prediction of bounding boxes with class labels within an end-to-end differentiable network. The YOLO model family treats object detection as a regression problem, utilizing CNN architectures like Darknet or Feature Pyramid Network (FPN) to extract features from input images. The output, passed to the detection head, predicts object coordinates and class probabilities. A non-maximum suppression (NMS) algorithm is then applied to eliminate redundant detections, enhancing overall model accuracy. The final output includes bounding boxes, class probabilities, and objectness scores for each detected object. Over time, YOLO models have evolved with improvements to the backbone network, detection head, and techniques like anchor boxes, resulting in increased accuracy and faster object detection.

4.2 YOLO V5

The YOLOv5 architecture, like its predecessors, is designed for real-time object detection. Here's a simplified explanation of the YOLOv5 architecture:



4.2.1 Backbone Network:

❖ YOLOv5 uses CSPDarknet53 as its backbone network. CSP stands for Cross-Stage Partial networks, a feature that enhances the flow of information between different stages of the network.

❖ CSPDarknet53 is a modified version of Darknet53, incorporating Cross-Stage Hierarchical Feature Aggregation (CHFA) for improved feature representation.

4.2.2 Neck:

❖ YOLOv5 introduces a PANet (Path Aggregation Network) as the neck. PANet helps in aggregating information from different scales and stages of the feature pyramid.

❖ The PANet allows the model to capture both fine-grained details and high-level semantic information, enhancing object detection across varying scales.

4.2.3 Head:

❖ The detection head consists of three prediction scales (small, medium, and large), each responsible for detecting objects at different scales in the image.

❖ Each scale predicts bounding box coordinates, objectness score, and class probabilities. The predictions are made at multiple scales to handle objects of different sizes.

4.2.4 Prediction Head Design:YOLOv5 utilizes the familiar anchor box mechanism for bounding box prediction. The model predicts the offsets for anchor boxes associated with each grid cell.The classification head outputs class probabilities using a sigmoid activation for each class independently. This allows the model to handle multi-class object detection efficiently.

4.2.5 Loss Function:YOLOv5 uses a combination of loss functions to train the model:

- ❖ Objectness Loss: Penalizes the model for misclassifying objectness (whether an object is present in the bounding box).
- ❖ Bounding Box Regression Loss: Measures the difference between predicted and true bounding box coordinates.
- ❖ Classification Loss: Penalizes the model for misclassifying object categories.

4.2.6 Training Strategy:

YOLOv5 typically undergoes a two-step training process: pre-training on a large dataset for general object detection tasks, followed by fine-tuning on a specific dataset for the target task.

Model Variants: YOLOv5 comes in different variants (s, m, l, x) that vary in terms of the number of parameters and computational complexity. Users can choose a model variant based on the trade-off between speed and accuracy that suits their specific application. Overall, YOLOv5 is designed for efficiency and effectiveness in object detection tasks.

4.3 YOLOV8

YOLOv8, the latest addition to the YOLO model family, offers a range of capabilities including object detection, image classification, and instance segmentation, Developed by Glenn Jocher, who also created the YOLOv5 model, YOLOv8 operates on a similar principle. It consists of a backbone, neck, and head, collectively known as the FNP. However, there are some alterations in the modules. In YOLOv8, the kernel size of the first convolutional layer in the backbone network and neck modules is 3x3, whereas it was 6x6 in YOLOv5. Additionally, all C3 modules in YOLOv5 are replaced by C2f in YOLOv8, following the YOLOv7 ELAN concept. YOLOv8 incorporates more skip connections and split operations, while removing two convolutional connection layers from the neck module. However, the most significant difference lies in the head module. YOLOv8 replaces the original coupling structure of YOLOv5 with a decoupling one. Another notable distinction is that YOLOv8 is an anchor-free model, whereas YOLOv5 is anchor-based. In an anchor-based model like YOLOv5, a predefined set of anchor boxes with different sizes and aspect ratios is used to predict the location and size of bounding boxes. These predictions are then adjusted based on the offset between the anchor boxes and the ground-truth boxes, enabling accurate detection of objects with varying sizes and aspect ratios. Conversely, an anchor-free model like YOLOv8 directly predicts the center point and size of bounding boxes, eliminating the need for anchor boxes and simplifying the model.

4.3.1 Backbone:

The updated CSPDarknet53 architecture, which uses a multi-scale approach by converting input features into five different scales labeled B1 through B5, is the central component of YOLOv8. The C2f module replaces the old Cross Stage Partial (CSP) module. It has pass-through connections to improve information flow and keep the structure lightweight. The output of the CBS module is produced by convolution on the input data, batch normalization, and SiLU activation. The Spatial Pyramid Pooling Fast (SPPF) module, which lowers computational cost and delay by sequentially connecting three max-pooling layers, is incorporated into the backbone to enable adaptive scaling of the output.

4.3.2 Neck:

YOLOv8 enhances feature fusion and localization through the incorporation of the PAN-FPN structure in its neck, inspired by the PANet architecture. Importantly, YOLOv8 efficiently balances model performance without compromising efficiency by simplifying the PAN structure and eliminating post-sampling convolution processes. The PAN-FPN configuration leverages two specific feature scales, P4-P5 and N4-N5, in the PAN and FPN structures, respectively. This arrangement ensures a comprehensive blend of deep semantic and shallow location information, employing a synergy of top-down and bottom-up approaches to ensure feature diversity and completeness.

4.3.3 Head:

YOLOv8's detection component uses a split-head architecture with distinct branches for bounding box regression prediction and object classification. Various loss functions are used on these branches: distributed focus loss (DFL), bounding box regression using CIOU, and classification using binary cross-entropy loss (BCE loss).

Model convergence is accelerated and detection accuracy is improved by this decoupled approach. The model assigns samples dynamically using a Task Specifier.

4.3.4 YOLOv8 Training

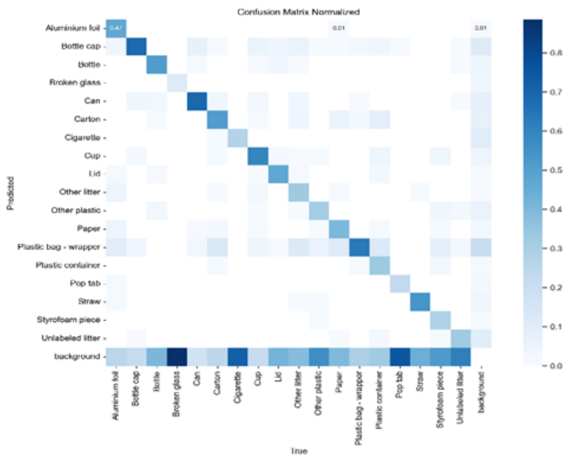
Initially, to conduct the training of the YOLOv8 model, the YOLOv8 repository is cloned to the computer. Furthermore, any additional dependencies and libraries are also installed to the standalone computer. Similar to YOLOv5, there are several versions of the YOLOv8 family model - YOLOv8s, v8m, v8l, v8x, out of which YOLOv8s is the smallest model and on the other end of the spectrum YOLOv8x is the largest one. For consistency purposes, as previously we trained YOLOv5's smallest model YOLOv5s and YOLOv5's medium model YOLOv5m, similarly in this also YOLOv8's smallest model (YOLOv8s) and YOLOv8's medium model (YOLOv8m) is utilized. Training, validation, Testing datasets divided into 70:20:10 ratio. The model is trained for 150 epochs on the same computer with the same specifications, where the batch size is chosen to be 16 training images. Once the model has completed training, a couple of test batch images, matrices and graphs are obtained. The training was completed in a total of 34.68 h. The immediate results show that the mean average precision is 0.927 across all classes.

5.RESULTS:

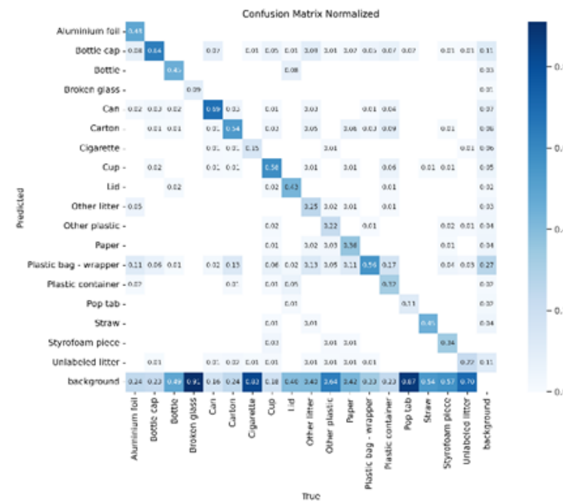
In our meticulous scrutiny of garbage detection models, the performance evaluation of garbage detection models is crucial for effective waste management systems. In our research, we compared two set of models having different version of YOLOv5 and YOLOv8 models (YOLOv5s Vs YOLOv8s) and (YOLOv5m Vs YOLOv8m) in terms of precision and mean Average Precision (mAP) to assess their suitability for identifying and localizing garbage objects in images.

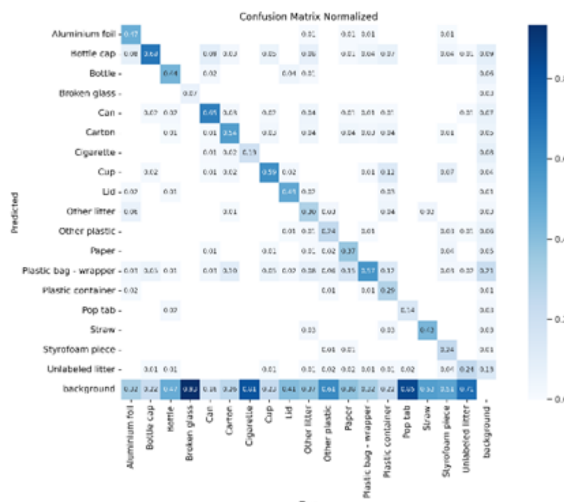
5.1 Confusion Matrices

A confusion matrix serves as a prevalent tool for evaluating the effectiveness of a classification model. It facilitates a comparison between the actual class labels and the predicted class labels derived from test images. This matrix proves instrumental in the computation of various performance metrics such as the F1 score, accuracy, precision, and recall. The accompanying figure delineates the confusion matrix for the YOLOv5s and YOLOv5m models, as well as the YOLOv8s and YOLOv8m models.



Confusion Matrix of YOLOv5 M and YOLOv8 M





Confusion Matrix of YOLOv5s and YOLOv8s

5.2 Precision-Confidence Curves

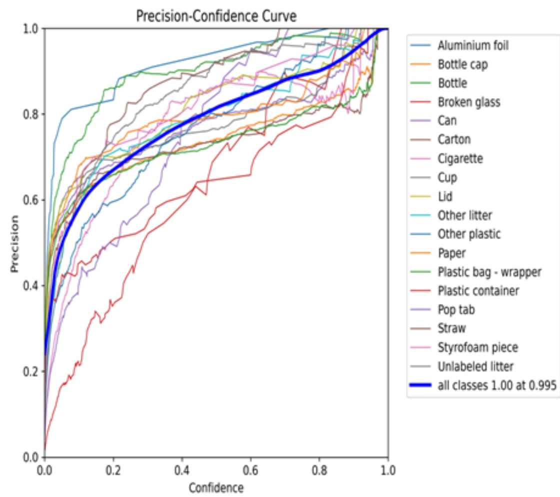
A precision-confidence curve is a graphical representation of the relationship between a classifier's confidence level and the precision of its predictions. It is frequently used in object identification and classification tasks to visualize the trade-off between a model's number of correct and incorrect predictions at different degrees of confidence. The curve represents the accuracy of the model's predictions at various levels of confidence. A high precision at a given threshold means that the classifier is making few erroneous positive predictions at that confidence level, but it may be missing some true positive predictions as well.

Based on the below figures, we can interpret the following

- Fig.(a) and (b), it can be concluded that at the same confidence level, YOLOv8m having precision 76.4% is more precise by 9.1% compared to the YOLOv5m having precision 67.3% model after running 150 epochs.
- Fig.(c) and (d), it can be concluded that at the same confidence level, YOLOv8s have precision 66.3% and are more precise by 0.9% compared to the YOLOv5s having precision 65.7% model after running 50 epochs.
- Classes like Aluminium foil, Bottle, Cigarette, Lid, other litter, Paper, straw giving best precision values more than 78%.

Table 1. Precision Comparison Table

	YOLOv5s	YOLOv8s	YOLOv5m	YOLOv8m
all	65.7	66.30	67.30	76.40
Aluminum foil	85.2	76.00	86.10	92.00
Bottle cap	63.7	69.70	69.80	71.10
Bottle	81.1	79.40	81.30	89.70
Broken glass	85	41.60	37.80	56.30
Can	71.8	71.50	68.80	74.80
Carton	54.9	68.50	64.50	70.90
Cigarette	57.2	63.90	69.60	82.90
Cup	69.6	69.40	69.10	74.30
Lid	76.9	73.30	69.60	78.90
Other litter	56	59.30	70.70	77.40
Other plastic	59.4	64.30	57.80	74.00
Paper	60.4	63.70	69.80	78.40
plastic bag - wrapper	58.9	67.70	67.10	70.70
plastic container	58.5	58.60	55.30	59.90
Pop tab	50.2	58.30	63.70	71.30
Straw	63.4	68.40	69.40	88.40
Styrofoam piece	69.8	76.70	70.40	78.80
Unlabeled litter	60.3	63.10	71.70	85.10



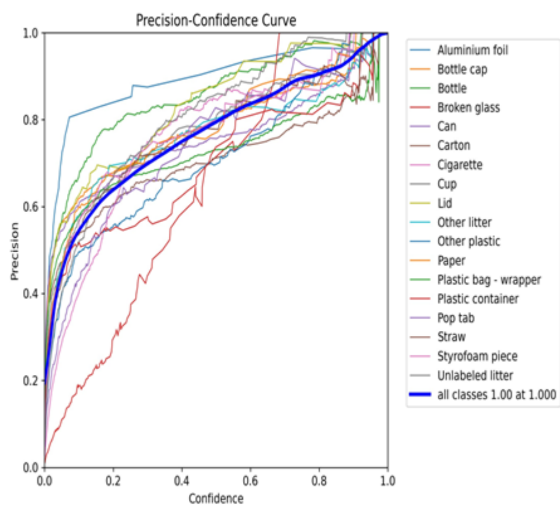


Fig a. P-C curve for YOLO V5m

Fig b. P-C curve fro YOLO V8m

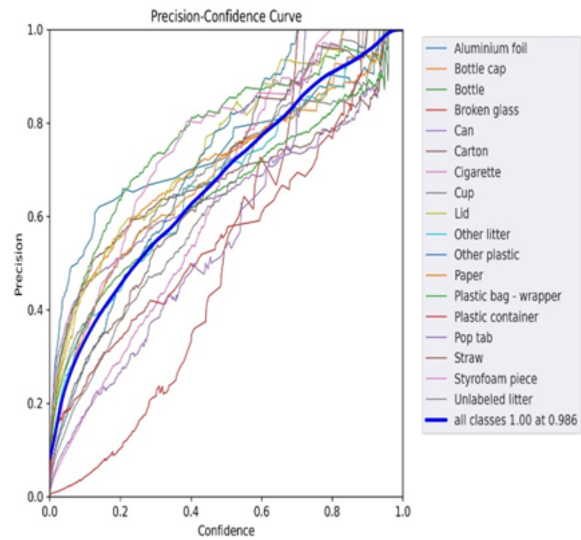


Fig c. P-C curve for YOLO V5s

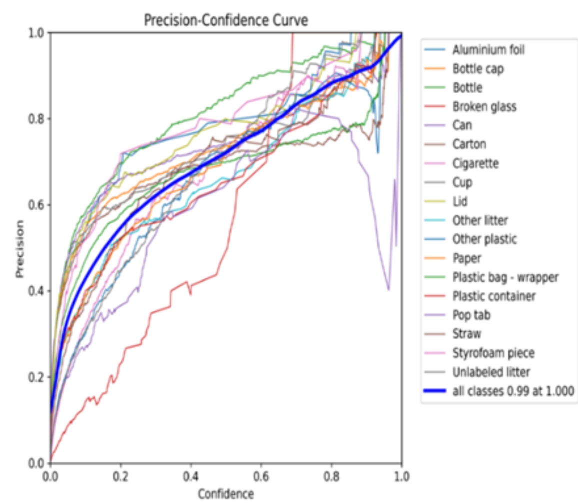


Fig d. P-C curve fro YOLO V8s

5.3 Comparison of Mean Average Precisions

Average precision (AP) is a popular criterion for evaluating object identification systems. It assesses the model's ability to recognize and forecast the location of items. The mAP value ranges from 0 to 1, with higher values indicating better performance. AP is a frequently used assessment metric in object detection that evaluates an algorithm's ability to detect and localize things of interest in an image or video. The AP value is calculated by averaging the precision values over a range of recall levels. Precision is the percentage of correct detections among all detected objects, while recall is the percentage of correct detections among all ground truth objects.

The Intersection over Union (IoU) threshold specifies the minimum amount of overlap between the expected result and ground-truth bounding boxes that must exist for a detection to be considered a true positive. When evaluating the effectiveness of an object detection model, the AP is often calculated at several IoU thresholds ranging from 0 to 1, with 0.5 being a popular choice. When only detections with IoU 0.5 are considered, the average accuracy (AP) at IoU = 0.5, or where the projected bounding box corresponds with the ground-truth bounding box by at least 50%, is determined. The mAP is an absolute metric that is not a function of confidence threshold.

In YOLO V8m, the mAP50 value reaches 48 whereas for YOLO V5m it is 47.5.

Table 2. mAP50 Comparison Table

	YOLO v5s	YOL Ov8s	YOLO v5m	YOLO v8m
all	41.4	41.10	47.50	48.00
Aluminum foil	53.7	48.40	50.40	54.10
Bottle cap	65.7	69.00	74.50	71.30
Bottle	56.6	52.10	59.60	61.00
Broken glass	8.99	7.00	10.40	11.80
Can	71	69.10	73.90	75.30
Carton	53.4	55.90	56.90	59.40
Cigarette	17.6	20.40	30.30	31.30
Cup	64.3	66.60	66.20	67.70
Lid	52.4	53.70	58.60	55.30
Other litter	27	32.30	37.70	35.00
Other plastic	24.1	27.00	34.40	35.80
Paper	38.8	38.00	46.60	48.60
plastic bag - wrapper	53.7	53.80	63.30	62.10
plastic container	38.1	34.40	40.30	36.90
Pop tab	13.4	15.10	25.00	25.20
Straw	43.8	41.20	53.50	56.70
Styrofoam piece	37.4	29.40	36.40	35.40
Unlabelled litter	25.9	26.30	37.10	39.90

In terms of mAP50 values, Classes like Bottle cap, Bottle, Can, Cup, plastic bag – wrapper are the Best performers, and, categories like Aluminium foil, Carton, Lid, Paper, Straw giving normal values, while Broken glass, Pop tab are worst classes

Table 3. mAP50:95 Comparison Table

	YOLO v5s	YOLO v8s	YOLO v5m	YOLO v8m
all	31.80	31.70	37.20	37.70
Aluminum foil	42.80	40.20	43.50	47.60
Bottle cap	51.70	53.50	62.10	59.00
Bottle	36.90	34.00	38.90	41.30
Broken glass	3.45	3.90	4.79	5.98
Can	58.40	54.70	61.80	62.50
Carton	44.30	47.20	49.70	53.20
Cigarette	8.19	9.78	15.00	14.40
Cup	52.30	55.50	57.70	56.60
Lid	42.60	45.00	48.00	46.60
Other litter	22.30	26.40	32.30	30.50
Other plastic	17.00	19.10	23.90	25.00
Paper	31.30	30.40	38.70	41.00
plastic bag - wrapper	39.70	40.10	50.30	49.80
plastic container	32.70	29.80	36.00	34.40
Pop tab	9.19	9.36	13.40	15.00
Straw	30.10	30.00	40.70	42.90
Styrofoam piece	33.10	26.00	29.80	29.20
Unlabeled litter	15.50	15.70	22.40	24.00

Overall, the YOLOv8m model has outperformed models like YOLOv5s, YOLOv8s, YOLOv5m for litter Detection

YOLOv8n exceeded Altinbas and Serif's model by 1.62%. Furthermore, the YOLOv8s has surpassed the YOLOv8n model by 2.24% and the YOLOv5 model by 3.32%.

The results of our experiments reveal a nuanced performance distinction, with YOLOv8s exhibiting a slightly higher precision compared to YOLOv5s, surpassing it by 0.6%. Interestingly, both models demonstrated an equivalent mean Average Precision (mAP@50), suggesting comparable accuracy in object localization. This finding emphasizes the fine-grained trade-offs between precision and mAP metrics in the context of garbage detection.

In our meticulous scrutiny of garbage detection models, a thorough comparative analysis was undertaken between YOLOv5m and YOLOv8m to discern their efficacy in identifying and localizing refuse objects across diverse environmental scenarios. The findings of our experiments delineate conspicuous distinctions in performance, with YOLOv8m exhibiting a noteworthy 9.1% increase in precision compared to YOLOv5m. Furthermore, YOLOv8m surpassed YOLOv5m by a marginal 0.5% in mean Average Precision (mAP), at IoU ≥ 0.5. A more profound exploration into specific use cases and potential optimization strategies for both YOLOv5m and YOLOv8m could furnish deeper insights into their respective efficacies and constraints in real-world waste management scenarios.

6. CONCLUSION AND FUTURE WORKS:

This research presents a pioneering application of YOLOv5m and YOLOv8m algorithms for the intelligent classification and management of urban waste, specifically leveraging the Taco dataset. Trained on a bespoke dataset of 2,452 images sourced from Taco waste recycling facilities, the neural network aimed to detect 28 specific waste types. Results showcase the efficacy of our approach in categorizing waste into five distinct groups, including bottles, cans, cardboard, and detergents, enabling nearly instantaneous waste detection. Comparative evaluations spanning YOLOv5m, YOLOv8m, and YOLOv8l models underscore the effectiveness of both YOLOv5m and YOLOv8m in waste classification. Despite the constrained size of the Taco dataset posing challenges for significant enhancements with YOLOv8m compared to YOLOv8s, future research aims to expand datasets to refine detection accuracy. The study recognizes the intricacies of detecting diverse waste images in the Taco dataset, opening avenues for further exploration of material-based waste classification. The proposed waste sorting strategy holds promise for enhancing waste disposal and recycling practices within the context of the Taco dataset. Future endeavors will prioritize optimising predictions and probabilities for various real-world wastes beyond the initial 28 classes, contributing to continuous advancements in waste management technology within the Taco dataset framework.

7. REFERENCES

[1] Chen H, Liu H, Sun T, Lou H, Duan X, Bi L, Liu L. MC-YOLOv5: A Multi-Class Small Object Detection Algorithm. Biomimetics (Basel). 2023 Aug 2;8(4):342. doi: 10.3390/biomimetics8040342. PMID: 37622947; PMCID: PMC10452785.

[2] Sun, Q.; Zhang, X.; Li, Y.; Wang, J. YOLOv5-OCDS: An Improved Garbage Detection Model Based on YOLOv5. *Electronics* 2023, *12*, 3403. https://doi.org/10.3390/electronics12163403

[3] Majchrowska, S., Mikołajczyk, A., Ferlin, M., Klawikowska, Z., Plantikow, M. A.,

Kwasigroch, A., & Majek, K. (2022). Deep learning-based waste detection in natural and urban environments. *Waste Management*, *138*, 274–284.

[4] Das, Dhrubajyoti & Kaushik, Deb & Sayeed, Taufique & Dhar, Pranab & Shimamura,Tetsuya. (2023). Outdoor Trash Detection in the Natural Environment Using a Deep Learning Model. IEEE Access. VOLUME11.97549-97566.10.1109/access.2023.3313166.

[5] Smoking behavior detection algorithm based on YOLOv8-MNC BY Zhong Wang1,2, Lanfang Lei1* and Peibei Shi2

[6] Object detection using yolo v5 BY L. Manikandan1 , Rapelli Malavika2 , Borlakunta Anjali3 , Sathuri Chandana

[7] Automatic Detection and Classification System of Domestic Waste via Multi-model Cascaded Convolutional Neural Network

[8] Skip-YOLO: Domestic Garbage Detection Using Deep Learning Method in Complex Multi-scenes | International Journal of Computational Intelligence Systems

[9] YOLO-Based Object Detection for Separate Collection of Recyclables and Capacity Monitoring of Trash Bins

[10]UAV-YOLOv8: A Small-Object-Detection Model Based on Improved YOLOv8 for UAV Aerial Photography Scenarios