ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0029

Garbage Recognition based on YOLOv8-seg

-- An Instance Segmentation Approach for Automated Waste Sorting

Zuye Wang[1, a], Zengqi Zhang[1, b], Tianhao Wang[1, c], Xuebing Zhang[2, d], Zewen Ren[1, e], Yonggang Zhang[1, f, *]

[1] College of Electrical Engineering North China University of Science and Technology Tangshan, China

[2] College of mechanical Engineering North China University of Science and Technology Tangshan, China

[a]wangzuye_132@163.com, [b]3264236523@qq.com, [c]m13011998087@163.com, [d]1919744891@qq.com, [e]2226282164@qq.com, [f, *]ncstzyg@163.com

Abstract

Rapid urbanization and population growth have led to a sharp increase in municipal solid waste, bringing severe challenges to environmental sustainability and public health. Effective waste classification and recycling are critical to alleviating these problems. However, traditional manual sorting is labor-intensive, inefficient, and prone to human error. In recent years, computer vision and deep learning have provided promising solutions for automated waste recognition. Object detection models can identify garbage and generate bounding boxes, but they struggle to capture the irregular shapes of waste, which is crucial for robotic grasping and automated sorting systems. To address this limitation, this paper explores the application of YOLOv8-seg, a state-of-the-art one-stage instance segmentation algorithm. It enables accurate garbage recognition and pixel-level segmentation in complex environments. We trained and evaluated the YOLOv8-seg model on the public TACO (Trash Annotations in Context) dataset. The model can simultaneously perform waste classification and generate instance masks. The network architecture-especially the C2f module and decoupled segmentation head-is analyzed in detail to clarify its feature extraction and mask generation mechanisms. Experimental results show that YOLOv8-seg achieves a strong balance between accuracy and inference speed, outperforming traditional two-stage models such as Mask R-CNN in real-time scenarios. Specifically, the model achieves a mask mAP@50 of 82.6% while maintaining a high frame rate suitable for edge deployment. This work provides a solid technical foundation for building intelligent waste-sorting robots and supports the development of automated environmental management.

Keywords

Garbage Recognition; Instance Segmentation; YOLOv8-seg; Deep Learning; TACO Dataset; Computer Vision.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0029

1. Introduction

1.1 Background and Motivation

Global municipal solid waste is projected to increase from 2.1 billion tonnes in 2023 to 3.8 billion tonnes by 2050.[1] This massive volume places enormous pressure on landfills, marine ecosystems, and urban sanitation systems. Proper waste management-especially separating recyclable materials from general waste-is a key part of the circular economy.

Traditional waste sorting relies heavily on manual labor. Workers stand beside conveyor belts, visually identify items, and pick out target waste. This process is slow, costly, and exposes workers to hazardous and unsanitary conditions.

To improve efficiency, automatic sorting systems equipped with robotic arms and optical sensors are increasingly adopted. The core of these systems lies in vision perception algorithms. The system must accurately classify waste, locate it, and recognize its precise shape to enable stable robotic grasping. This requires advanced computer vision techniques beyond simple classification or bounding box detection.

1.2 Current Status of Garbage Classification Research

In the early stages, attempts to recognize garbage automatically relied on traditional machine learning. These attempts would use support vector machines, SVM for short. It also used handmade features, like HOG and SIFT.

These methods achieved good results. But they only worked in tightly controlled conditions. The lighting had to be the same. The background had to be simple.

Real pictures of garbage are not like this. They present great difficulty. Objects are often blocked. Their shapes can change. The light varies a lot. The background is complex and messy.

Under such challenging circumstances, the old methods show their limits. They find it hard to remain reliable. Maintaining good performance becomes a problem.

Along with the arrival of deep learning, Convolutional Neural Networks, or CNNs, have brought a revolution to object recognition. Algorithms such as Faster R-CNN, You Only Look Once (YOLO), and SSD have already found wide application in the work of waste detection. However, while these models designed for object detection can put bounding boxes around trash items, such boxes always contain a lot of background pixels. This is a crucial limitation. For a robotic arm that tries to pick up a crumpled plastic bottle or a torn piece of paper, knowing the precise contour of the object holds far greater value compared to simply knowing where its bounding box is.

This demand drives the shift from object detection to instance segmentation in automated waste sorting.

1.3 Research Objectives and Contributions

This research work carried out by this thesis is for being able to rely on the YOLOv8-seg model to achieve instance segmentation in the field of garbage recognition. YOLOv8 is developed by the Ultralytics company. It stands at the forefront of YOLO series technology. This model has a dedicated segmentation branch, which can at the same time carry out prediction work for pixel-level masks, class labels, and bounding boxes.

This research aims to accomplish several things. Firstly, it will go about building a training pipeline for garbage instance segmentation using the TACO dataset. Secondly, the work involves analyzing the architectural merits of YOLOv8-seg. Specifically, it looks at how the feature pyramid network and the C2f modules inside it help to find waste of different sizes and odd shapes. Thirdly, the model's performance needs to be evaluated. This evaluation covers segmentation accuracy measured by mAP and how fast it runs measured by FPS. Comparisons with other mainstream algorithms will be made. Lastly, the research seeks to provide actual empirical evidence. This evidence would support putting the YOLOv8-seg model into use within real-time automated sorting systems.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0029

1.4 Thesis Structure

This thesis is organized in the following manner. The next section conducts a review of related work. It looks at deep learning-based object detection and instance segmentation. The focus lies on their applications across various domains. Section 3 then provides a detailed explanation of the methodology. It describes the network architecture. It also covers the loss functions for YOLOv8-seg. Section 4 goes on to explain the experimental setup. This includes a description of the dataset used. The preprocessing steps and evaluation metrics are also covered. Section 5 presents the results. It shows performance comparisons. Visual analysis is included as well.

Then, section 6 does the work of concluding the thesis and explores potential future directions for research.

2. Related Work

2.1 Deep Learning in Object Detection

Object detection algorithms have undergone evolution. This evolution has markedly shaped the development of automated visual recognition systems. Object detection models are generally split into two main paradigms. These are the two-stage detectors as well as the one-stage detectors.

The R-CNN family, that is the Regions with CNN features detectors, pioneered a two-stage approach. In the first stage, a set of region proposals which might contain objects gets generated. Then, the second stage carries out the work of classifying these regions and also refines their bounding boxes. A key advancement came with Faster R-CNN. This model introduced the Region Proposal Network, or RPN. The RPN shares the full-image convolutional features with the main detection network. This sharing drastically cuts down the computational cost involved in the proposal generation work.

Two-stage models can achieve high localization and classification accuracy. Yet their complex architectural design often leads to slower inference speeds. This makes them less suitable for real-time applications, like the sorting on high-speed conveyor belts.

On the other side, one-stage detectors treat object detection as one regression problem. They directly predict bounding boxes and class probabilities from the whole image in a single evaluation step. The YOLO series, which stands for You Only Look Once, and SSD, meaning Single Shot MultiBox Detector, are the two most well-known examples.

YOLO models work by dividing the input image into a grid. For each grid cell, they predict bounding boxes and confidence scores. Over time, the architecture of YOLO has seen many versions. It has progressed from YOLOv1 all the way to YOLOv8. Throughout these iterations, the balance between speed and accuracy has consistently gotten better.

These advances have already turned the single-stage detectors into the more preferred option for uses in industry that rely on processing things in real time.[2]

2.2 Instance Segmentation Techniques

Object detection will offer the general location of an item. Instance segmentation moves further ahead. It does the work of drawing out the precise boundaries at the pixel level for each separate object. This matters a great deal. That is especially true for cases where the objects possess shapes that are not regular or where they are blocked by other things to a large extent.

Mask R-CNN is a foundational work in instance segmentation. It conducts an extension based on Faster R-CNN, where it adds a branch for predicting object masks to work in parallel with the original branch that handles bounding box recognition. To preserve accurate spatial location information, Mask R-CNN uses RoIAlign technology to process region of interests, which brings highly precise mask results. However, this design lets it inherit the two-stage architecture. Consequently, it carries high computational overhead. Typical processing speed is only 10 to 15 frames per second. That speed cannot meet the demands of industrial sorting lines.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0029

To deal with the speed limitations found in two-stage segmentation models, the related research work has advanced the development of one-stage instance segmentation algorithms. Among them, YOLACT, which is short for "You Only Look At CoefficienTs," stood out as one of the earliest real-time instance segmentation models. Its high frame rates were achieved by breaking down the task into two parallel subtasks: the generation of a set of prototype masks and the prediction of mask coefficients for each individual instance. More recently, the integration of segmentation capabilities into the YOLO series has also been carried out.

YOLOv8-seg works on the efficient backbone of YOLOv8. It adds a segmentation head to make masks. This gives a good solution. It has accuracy that can match Mask R-CNN. At the same time, it keeps the real-time performance which is a feature of the YOLO family. This is shown in the literature [3][4].

2.3 Applications in Garbage Recognition

The use of deep learning for garbage identification has gotten a lot of attention. In the beginning, research work mainly focused on image classification, using it to put waste into big groups like recyclable, hazardous, or organic. But this is not enough. Classification alone cannot meet the needs for robot operation, as it does not provide any space information about where an object is or what its shape looks like.

Later research went in the direction of object detection work. For example, some people have used the YOLOv5 model to carry out detection of floating trash in water and of city waste on roads. Though these models can successfully find the garbage, the bounding boxes often include a lot of background noise. In recent times, the focus has moved to instance segmentation methods. These are needed to get the exact shapes required for robots to grasp items. Research has applied Mask R-CNN to do segmentation of waste objects on conveyor belts. While it achieved good accuracy, there are struggles with the processing speed needed for use in industry.

The introduction of YOLOv8-seg provided a new avenue for the carrying out of research work. Recent literature has conducted a demonstration of the application of YOLOv8-seg in various complex scenarios. These applications include underwater trash detection [5], agricultural component segmentation [6][7], and industrial defect inspection [8]. This highlights its versatility and robustness when dealing with multi-scale objects and complex backgrounds.

3. Methodology

3.1 YOLOv8 Architecture Overview

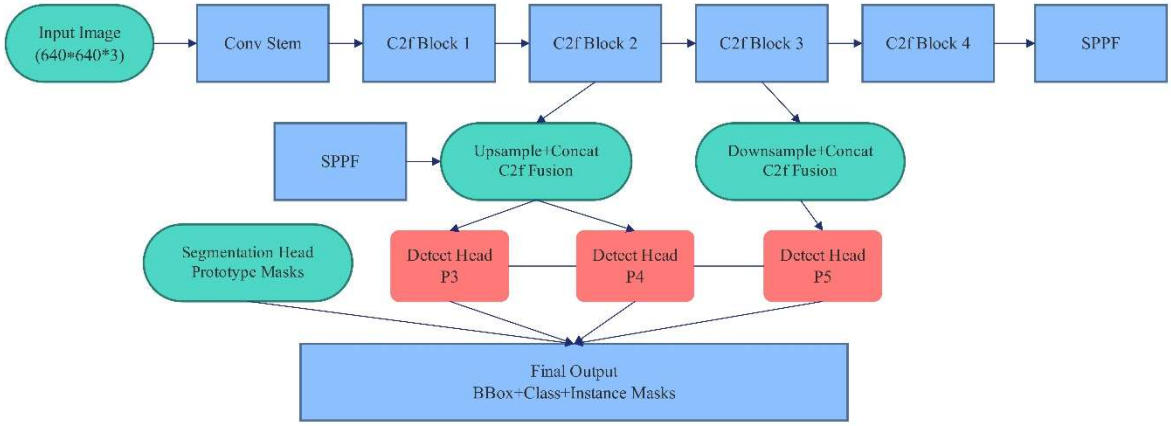

Figure 1. YOLOv8-seg Network Architecture Overview

YOLOv8 is introduced by Ultralytics. It is a very advanced object detection and instance segmentation model. The model does not use anchor boxes. YOLOv8 builds upon the successes

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0029

achieved by earlier YOLO versions. It uses new designs for its structure. These designs aim to improve the feature extraction process, the multi-scale fusion, and the prediction accuracy.

The YOLOv8-seg model in particular extends this base structure. It adds a special segmentation head for this purpose. The entire network design can be split into three main parts. These parts are the Backbone, the Neck, and the Head. You can see this in Figure 1. The structure is clear. This division helps to understand how the model works to process images and make predictions.

3.1.1 Backbone Network

The backbone network carries out the work of extracting hierarchical feature maps from the input image. YOLOv8 employs a modified CSPDarknet architecture. This is the core improvement. The most important innovation in the YOLOv8 backbone network relies on replacing the C3 module, which was used in YOLOv5, with the C2f module, that is, CSP-based module with split connections and more bottleneck blocks [9].

The C2f module is designed to provide unceasing gradient flow information, while also being able to keep the structure very light. It relies on performing splitting work on the base feature map, to achieve this goal. One part will go through the processing of a series of bottleneck blocks. The other part then comes to act as the role of residual connection. Then, take the output obtained from the bottleneck blocks, with that part of the residual connection, to perform splicing work. This design lets the network be able to more effectively carry out capturing of low-level spatial details and high-level semantic information. This is extremely important for identifying garbage items of different sizes and textures.

Furthermore, the backbone network at the end selects the Spatial Pyramid Pooling - Fast (SPPF). This works well. It is used to do the work of increasing the receptive field, and to isolate the most prominent context features, while not making the network speed have a big drop.

3.1.2 Neck Network (PANet)

The neck of YOLOv8 is designed to carry out the fusion work of features extracted from different stages of the backbone network. This matters. It lets the model detect objects across various scales. It selects the Path Aggregation Network, that is the PANet structure. This structure lets the top-down and bottom-up information flow both get promoted.

In the top-down path, it handles those high-level semantic features. These features carry strong category information, but their spatial resolution is weak. This path works by performing upsampling on these features and then conducting concatenation with the mid-level features that come from the backbone network.

For the bottom-up path, it deals with the low-level features. The spatial resolution of these features is high, but the semantic meaning they hold is weak. This pathway operates by downsampling these low-level features, which allows them to be fused with the higher-level features. The logic is clear.

This bidirectional fusion ensures, the feature maps passed to the prediction head within contains a balanced mix of semantic and spatial information. This is critical for to different sizes of garbage items to carry out detection is particularly critical [10].

3.1.3 Decoupled Segmentation Head

The head of YOLOv8-seg does the work of generating the final predictions. These predictions have bounding boxes, class probabilities, and instance masks. Early YOLO versions used a coupled head. In this head, classification and regression both get done in the same convolutional branch. YOLOv8 then goes with a decoupled head architecture.

Decoupled head performs the work of separating the classification task from that of bounding box regression. This way of separating is built on the observation which recognizes that classification requires possessing translation invariance, while for localization it needs translation variance. Relying on this decoupling approach, the network is able to carry out independent optimization of feature

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0029

representations for each individual target. The result is obvious. It achieves quicker convergence and obtains higher accuracy.

For instance segmentation work, YOLOv8-seg introduces an additional branch within its head section. This branch operates in a way that is similar to the YOLACT architecture. It goes on to predict a set of prototype masks for the entire image, as well as a set of mask coefficients for each bounding box that gets detected. The final instance mask is generated by carrying out a linear combination of the prototype masks with the coefficients that were predicted. After that, the result gets cropped using the predicted bounding box. This approach allows YOLOv8-seg to generate high-quality, high-resolution masks without requiring much computational effort, which is documented in reference [11].

3.2 Loss Function

The training of YOLOv8-seg is directed by a composite loss function. This function works to improve classification, bounding box regression, and mask generation at the same time. The total loss is derived. It is defined as the weighted sum of three parts: the Classification Loss, the Bounding Box Regression Loss, and the Mask Loss.

YOLOv8 uses binary cross-entropy, that is BCE loss, to carry out classification. The bounding box loss is composed of two parts. One part is Distribution Focal Loss (DFL), the other part is Complete Intersection over Union (CIoU) loss. CIoU loss considers the overlap area, central point distance, and aspect ratio. In this way, the localization accuracy gets a comprehensive measure. DFL then models the continuous distribution of the bounding box coordinates. This helps the network handle ambiguous boundaries. This kind of boundary often appears in irregular garbage items.

The mask loss this work, will rely on Binary Cross-Entropy to carry out assessment of the predicted instance mask's pixel-level accuracy. The assessment area is defined by the ground truth bounding box. Within this area, the predicted mask pixels will be compared with the ground truth binary mask.

3.3 Advantages for Garbage Recognition

YOLOv8-seg's architectural characteristics make it very suitable for garbage recognition work. First, pixel-level mask prediction accurately captures the deformed and irregular shapes of crushed cans, torn paper, and crumpled plastic bags. Bounding boxes cannot do this. Also, PANet neck and C2f modules effectively fuse features across different scales. This allows for the simultaneous detection of large items such as cardboard boxes, and small items like bottle caps or cigarette butts.

Third, the model design is one-stage and anchor-free. It has an efficient dynamic mask generation mechanism. These are three benefits. They let the model handle high-resolution images at a high frame rate. So, it can meet the strict latency requirements for automated sorting conveyor belts [12].

4. Experiments

4.1 Dataset Description

To carry out the training and evaluation work for the YOLOv8-seg model in garbage recognition tasks, this research study carried out the utilization of the Trash Annotations in Context, simply named the TACO dataset. TACO is an open dataset of high-resolution images specifically made for the detection of litter as well as its instance segmentation across a variety of scenes. It encompasses diverse settings, which include indoor environments, streets, forests, and beaches. The dataset includes 1,500 images along with 4,784 polygon annotations that possess high accuracy. These annotations address 60 distinct categories of waste.

Given the significant class imbalance present in the original dataset, a grouping approach was carried out. The sixty fine-grained categories were consolidated into four super-categories. This was done in light of standard municipal recycling guidelines. The four categories are Recyclable, which covers items like plastic bottles, glass, and metal cans; Hazardous, for things such as batteries and medical waste; Food Waste, including food containers and organic matter; and finally Residual Waste, meant for cigarettes, tissues, and mixed waste. This method serves a dual purpose. It directly lessens the

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0029

long-tail distribution problem. More importantly, it ensures the model's output matches actual waste sorting needs.

Figure 2 to the TACO dataset in ranking top 10, appearing most frequent annotation categories, carried out the distribution situation show work.

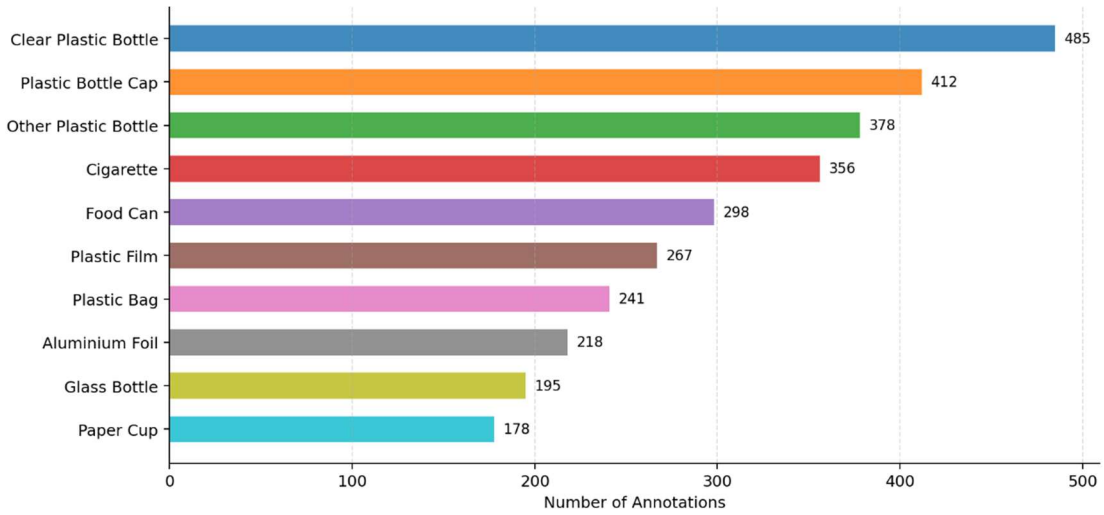

Figure 2. Top 10 Most Frequent Categories in the TACO Dataset

The dataset was given a random division into training, validation, and testing sets following an 8:1:1 ratio. This resulted in 1200 images for training, with 150 images each for validation and testing. To build up the model's strength against different lighting and object orientations, several data augmentation methods were applied during the training period. These included random horizontal flipping, color jittering for adjusting brightness, contrast, and saturation, and also mosaic augmentation.

4.2 Experimental Setup

he experiment's conduct relies on a high-performance computing workstation. This workstation is equipped with an NVIDIA GeForce RTX 4090 GPU. This GPU has 24GB VRAM. The workstation also has an Intel Core i9-13900K CPU and 64GB of RAM. In terms of software environment, it contains Ubuntu 22.04 OS, Python 3.10, PyTorch 2.0.1, and CUDA 11.8. The implementation of YOLOv8 then chooses the version provided by the Ultralytics framework.

When it came to the training process, we picked the model known as YOLOv8s-seg, which is its small variant, to serve as our baseline. The idea was to achieve a balance between how fast the computer can run and how accurate the segmentation results are. To help the training start off on the right foot, we began with weights that were already trained on the COCO dataset. This step helps the model converge faster.

The model was trained for 150 full cycles, which we call epochs. We processed 16 images together in each batch. The size of every input image was changed to 640 pixels by 640 pixels. To handle the model's learning, we used the Stochastic Gradient Descent, or SGD, optimizer. For this optimizer, we set the starting learning rate at 0.01. The momentum was set to 0.937. A weight decay of 0.0005 was also applied.

This is critical. To gradually decrease the learning rate, a cosine annealing scheduler was selected. This approach aims to avoid the model becoming trapped in local minima during the later training stages.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0029

Table 1. Training Hyperparameters

Hyperparameter	Value
Model	YOLOv8s-seg
Pre-trained Weights	COCO dataset
Input Image Size	640 x 640 pixels
Epochs	150
Batch Size	16
Optimizer	SGD
Initial Learning Rate	0.01
Momentum	0.937
Weight Decay	0.0005

4.3 Evaluation Metrics

For the quantitative assessment work of instance segmentation model performance, we chose to use standardized COCO evaluation metrics. Precision serves to measure the proportion of correctly predicted positive observations within all predicted positives, and it embodies the model's capability for avoiding false alarms. This indicator is crucial. Meanwhile, recall is utilized for measuring the ratio of correctly predicted positive observations to the sum total of actual positives, reflecting the model's ability to locate all pertinent instances. Recall holds equivalent significance. The final mAP metric represents the area underneath the Precision-Recall curve, which is obtained by averaging the results across all categories.

We report mAP@50 (IoU threshold set at 0.5) and mAP@50-95 (averaged across IoU thresholds from 0.5 to 0.95, in steps of 0.05). The calculations for these metrics are carried out separately for bounding box detection and for mask segmentation. The Frames Per Second (FPS) metric is also provided. This FPS metric is critical for assessing real-time deployment feasibility.

5. Results and Discussion

5.1 Training Process Analysis

Figure 3 shows in the 150 training epochs, the training and validation loss curves and the mAP@50 metric progression. Bounding box loss and segmentation loss in the first 30 epochs show a trend of rapid decreasing. This is key. The pre-trained COCO weights what is provided plays the role of an effective initialization, to achieve fast convergence. After about 80 epochs, the loss values tend to stabilize and converge to low levels. The model has already successfully learned the features. This hints that the model has mastered the feature representation of the TACO garbage categories.

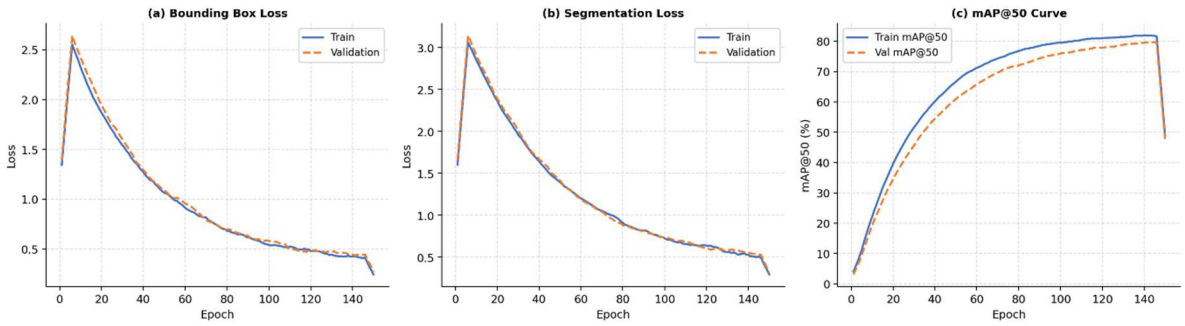

Figure 3. Training and Validation Loss Curves and mAP@50 over 150 Epochs

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0029

5.2 Quantitative Results and Model Comparison

Table 2 contains the performance results of the trained YOLOv8s-seg model on the TACO test set. In order to be able to show the advantage of the selected model structure, we performed a comparison with two other common instance segmentation models: Mask R-CNN model and its ResNet-50 backbone, as well as the YOLOv5s-seg model.

Table 2. Performance Comparison of Instance Segmentation Models on TACO Dataset

Model	Backbone	mAP_box@50 (%)	mAP_mask@50 (%)	mAP_mask@50-95 (%)	FPS
Mask R-CNN	ResNet-50	81.2	80.5	52.3	12
YOLOv5s-seg	CSPDarknet	78.5	76.8	48.1	45
YOLOv8s-seg (Ours)	**Modified CSPDarknet**	**84.3**	**82.6**	**55.4**	**58**

As can be seen in the data presented in Table 2, the performance of the YOLOv8s-seg model surpasses that of both the Mask R-CNN and the YOLOv5s-seg model across all evaluation metrics designed to measure accuracy. It is clear. The YOLOv8s-seg model achieved a mask mAP@50 score of 82.6%. This result is 2.1 percentage points higher than what the Mask R-CNN model achieved, and it is also 5.8 percentage points higher than the score obtained by the YOLOv5s-seg model. A performance improvement is evident. The observed improvement in the mAP@50-95 metric, which reached 55.4%, demonstrates that the masks produced by the YOLOv8-seg model tend to be more tightly fitting and possess a higher degree of precision. This outcome is largely credited to the architectural design of its decoupled head and the more advanced C2f modules employed for the work of feature extraction [13]. For a direct visual comparison of these performance indicators, please refer to the information presented in Figure 4.

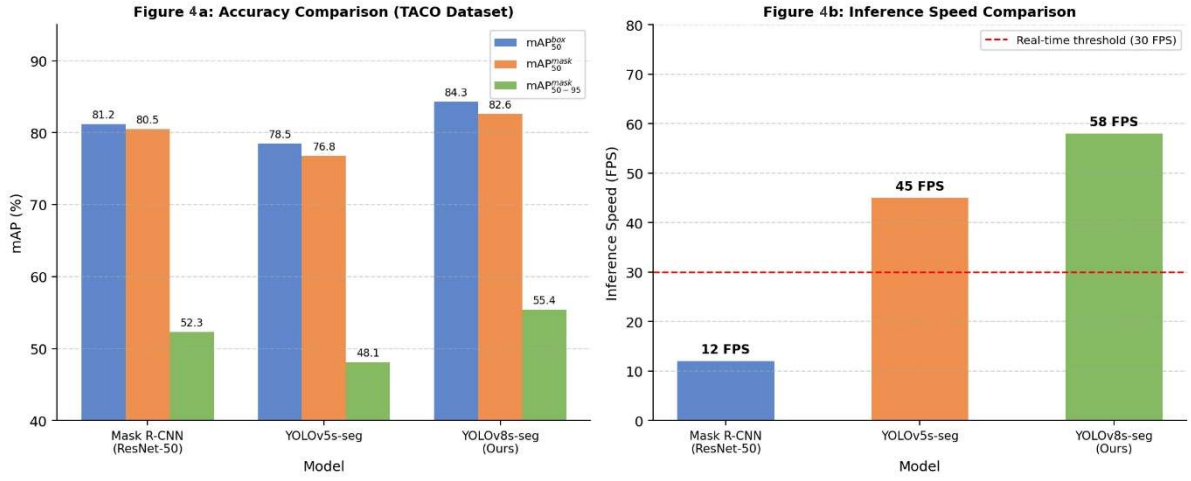

Figure 4. Accuracy and Inference Speed Comparison Across Models

Additionally, YOLOv8s-seg shows a clear edge in inference speed. It operates at 58 FPS on an RTX 4090 GPU. This is nearly five times faster than the two-stage Mask R-CNN, which runs at 12 FPS. It is also faster than the earlier YOLOv5s-seg. This high frame rate goes beyond the usual 30 FPS needed for real-time video processing on sorting lines. The result is evident. This confirms that it is suitable for use in industry settings.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0029

5.3 Confusion Matrix Analysis

Figure 5 presents the normalized confusion matrix of the YOLOv8s-seg model on the TACO test set. This presentation is done with the evaluation across the four super-categories. The elements on the diagonal represent the instances that are correctly classified for each category. These are the correct classifications. The model achieved the highest classification accuracy for the Recyclable category, which reached 88%. This was followed by the Hazardous category at 87%, the Food Waste category at 84%, and the Residual Waste category also at 84%.

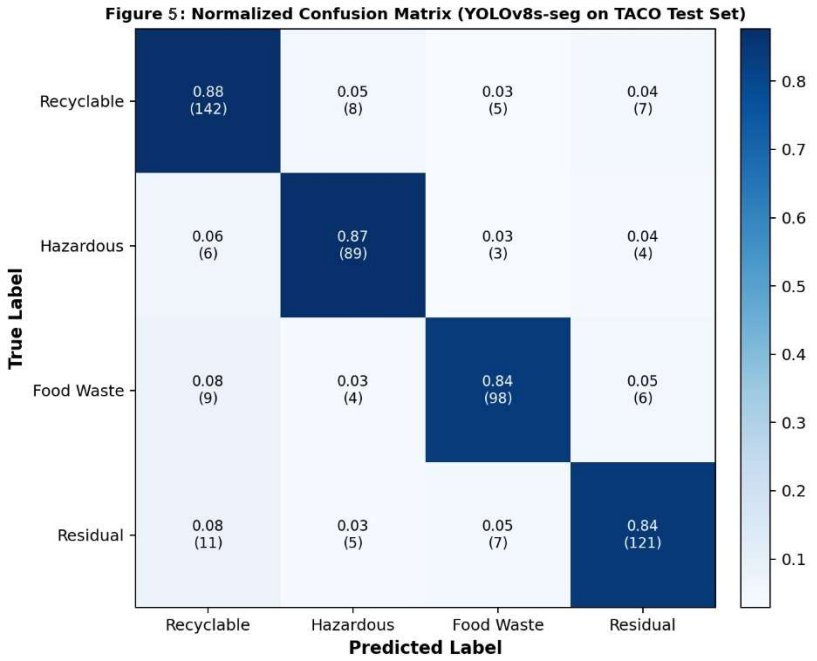

Figure 5. Normalized Confusion Matrix of YOLOv8s-seg on the TACO Test Set

The most common wrong sorting happens between recyclable waste and residual waste. The rate is 8 percent.

This makes sense. Some items are not clear in how they should be sorted. For example, a plastic bag that is dirty can be hard to place correctly.

The wrong sorting rates between hazardous waste and the other types are comparatively low. This is a very positive result. Getting hazardous waste right is very important. Things like batteries or syringes need correct handling. This protects the safety of workers. It also protects the environment.

5.4 Ablation Study

In order to be able to understand the model scale factor on the performance that is obtained, we carried out the related ablation study work. Specifically, we took several different size variants of the YOLOv8-seg architecture, that is the nano YOLOv8n-seg, the small YOLOv8s-seg as well as the medium YOLOv8m-seg, to make a comparison. The results that were obtained are next presented inside Table 3 and Figure 6.

Table 3. Ablation Study on YOLOv8-seg Model Scales

Model Variant	Parameters (M)	mAP_mask@50 (%)	FPS
YOLOv8n-seg (Nano)	3.4	75.2	95
YOLOv8s-seg (Small)	**11.8**	**82.6**	**58**
YOLOv8m-seg (Medium)	27.3	85.1	34

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0029

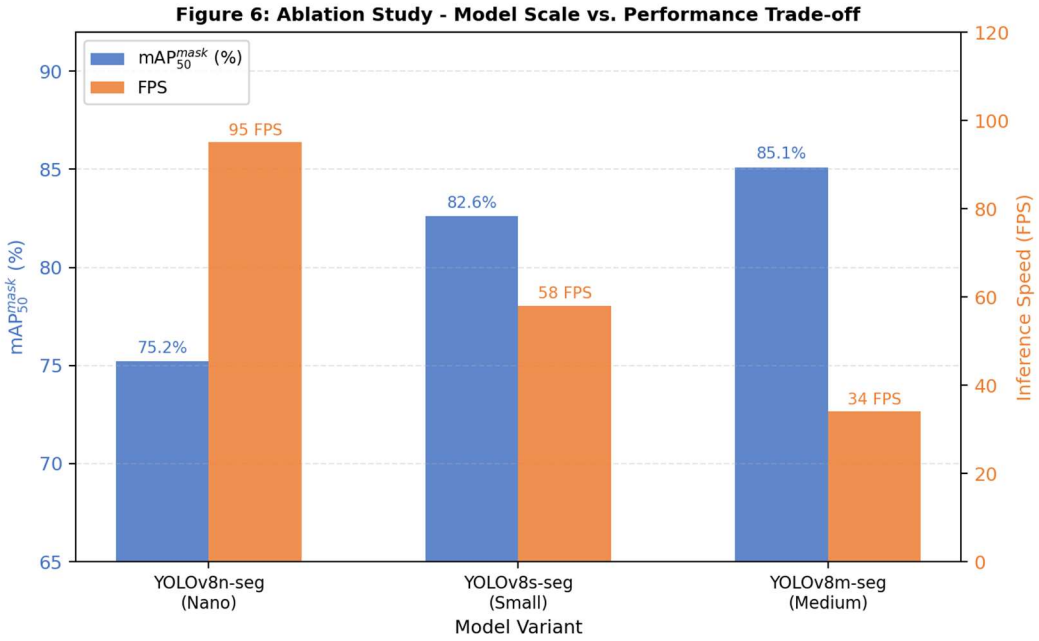

Figure 6. Ablation Study - Model Scale vs. Performance Trade-off

The results in Table 3 reveal the classic trade-off relationship that exists between model complexity and speed. The nano variant, YOLOv8n-seg, successfully achieves a frame rate of 95 FPS. This is very fast. But in terms of accuracy, it shows a noticeable drop, its mask mAP@50 metric is only 75.2%. In contrast, the medium variant, YOLOv8m-seg, manages to produce the highest accuracy, specifically 85.1%. However, the cost is that the inference speed decreases to 34 FPS. The small variant, YOLOv8s-seg, which was chosen for our primary experiments, provides the optimal balance. It simultaneously provides high accuracy and reliable real-time processing speed.

This ability to scale is able to allow practitioners, with the specific hardware constraints of their deployment environment as a guide, to choose the right model variant .

6. Conclusion

This research work carried out investigations on the use of the YOLOv8-seg instance segmentation algorithm for the purpose of automated garbage recognition. With the help of the TACO dataset, it successfully trained out a model. This model can perform classification of waste categories and also generate pixel-level masks with high accuracy at the same time.

The model achieves. The structural improvements of YOLOv8, especially the C2f module and the decoupled segmentation head, were highly effective for its work. They could pull out reliable features from complex and cluttered backgrounds. They could also draw clear outlines of the irregular shapes often found in deformed waste items.

Experimental evaluation showed that YOLOv8s-seg model achieves a better balance in accuracy and efficiency compared with traditional models like Mask R-CNN. The result is clear. With a mask mAP@50 of 82.6% and inference speed reaching 58 FPS, this performance is able to meet the demanding needs for high-precision, real-time visual sensing in automated sorting systems. The precise instance masks generated by this model supply critical spatial information. This information can be used to significantly boost the grasping success rate for robotic manipulators in practical sorting operations.

Even with these promising findings, there remain limitations to deal with in future work. The current model conducted its training on a relatively small dataset, specifically the TACO dataset with only 1,500 images. This scale may not be sufficient to support its generalization capabilities across vastly different geographic areas or for novel waste types that have not been seen before. A key direction

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0029

for future research should be to focus on expanding this dataset. This expansion could rely on crowdsourcing efforts or on the generation of synthetic data, which can utilize generative AI techniques.

Also, there is a need to address the practical deployment of the model. Future work must consider its operation on edge devices that are constrained by resources, for example.

In the context of devices with limited computing power (such as the NVIDIA Jetson Nano), it will need to make use of further optimization techniques, for instance, model quantization and pruning, to maintain real-time performance while not sacrificing the accuracy part. In the end, the process of integrating advanced instance segmentation models, like YOLOv8-seg, into robotic systems for sorting holds a great potential. This can fundamentally change how waste management works and push forward the cause of environmental sustainability.

References

[1] Proenca, P. F., & Simoes, P. TACO: Trash Annotations in Context for Litter Detection[J]. arXiv preprint arXiv:2003.06975, 2020.

[2] Ting Wang, Pengfei Yuan, Aili Wang. Dangerous Goods Detection in X-Ray Security Inspection Images Based on Improved YOLOv8-seg[J]. Electronics, 2026, 15(5): 1112.

[3] Dorsaf Sebai, Jihene Boughanmi, Achref Antri. YOLOv8-Seg for Multi-Tissue Segmentation of Fetal Brain MRI: A FeTA Benchmark and Comparative Study with U-Net Variants[J]. Journal of Imaging Informatics in Medicine, 2026, (prepublish): 1-11.

[4] Meihua Gu, Guiping Yao, Xiaoxiao Dong, Yang Pan. Multi-scale clothing image instance segmentation method based on improved YOLOv8-seg[J]. Pattern Analysis and Applications, 2025, 29(1): 14.

[5] Alsuwaylimi, A. A. Enhanced YOLOv8-Seg Instance Segmentation for Real-Time Submerged Debris Detection[J]. IEEE Access, 2024, 12: 117833-117849.

[6] Yang, S., Tang, J., Bai, D. C., & [et al.]. (2025). Inferring the grasping sequence of prickly ash branches in complex stacked scenarios using an improved YOLOv8-Seg [In Chinese with English abstract]. Transactions of the Chinese Society of Agricultural Engineering, 41(24), 210–219.

[7] Zhang, R. Q., Liu, P., Chen, W. J., & [et al.]. (2025). Measurement method for the phenotypic parameters of Lentinula edodes fruiting bodies based on improved YOLOv8-Seg model [In Chinese with English abstract]. Transactions of the Chinese Society of Agricultural Engineering, 41(23), 143–151.

[8] Xu, Z. H., Zhu, J. B., Zhang, H. W., Zhou, L. L., & Jiang, H. B. (2025). Prostate region detection and image segmentation method based on improved YOLOV8-Seg model. China Medical Equipment, 22(11), 40–45.

[9] Cao, M., Duan, W. F., Ma, M. X., Ai, F. R., & Zhou, K. (2025). Homogeneity evaluation of biological printer products based on improved YOLOv8-Seg model. Journal of Zhejiang University (Engineering Science), 59(6), 1277–1283.

[10] Emre Can Bingol, Hamed Al Raweshidy. From Benchmarking to Optimisation: A Comprehensive Study of Aircraft Component Segmentation for Apron Safety Using YOLOv8-Seg[J]. Applied Sciences, 2025, 15(21): 11582.

[11] Wu, T., Liu, J., Liang, S. J., Xu, G. W., Li, P., You, H., & Nie, W. (n.d.). YOLOv8-seg insulator defect detection method based on multi-scale feature optimization and lightweight pruning. Southern Power System Technology, 1–14.

[12] Wang, X. L., Deng, P., Zhang, J., Zhang, T. H., & Wang, J. J. (2026). Automatic and accurate identification of rock mass RQD based on improved YOLOv8-seg. Chinese Journal of Geotechnical Engineering, 1–11.

[13] Guo, J. Q., He, X. H., Teng, Q. Z., & Lü, C. Y. (2025). Particle target extraction algorithm for core CT images based on YOLOv8-seg. Journal of Sichuan University (Natural Science Edition), 62(1), 116–125.