Enhancing Small Object Detection in Waste Management: A Modified YOLOv8 Architecture with P2 High-Resolution Head

M. Ervin
Department of Information Systems
Hasanuddin University
Makassar, Indonesia
ervinm23h@student.unhas.ac.id

***Abstract*—The rapid accumulation of waste requires efficient automated sorting systems. However, standard object detection models often fail to identify small, crushed, or overlapping waste items, leading to significant missed detections (low recall). This study proposes YOLOv8s-P2, a modified architecture integrating a high-resolution P2 detection head (Stride 4) to enhance sensitivity toward minute and deformed targets. Using a challenging dataset from Roboflow containing four waste classes, the model was trained to maximize feature extraction from low-level backbone layers. Experimental results demonstrate exceptional performance, achieving a Mean Average Precision (mAP50) of 97.8% and a Recall of 94.3%. Notably, the model achieved perfect recall (100%) for difficult classes such as crushed inorganic bottles and organic peels. These findings confirm that the P2-Head modification effectively overcomes occlusion challenges, offering a robust solution for real-time waste classification.**

***Index Terms*—YOLOv8, waste detection, small object detection, P2 head, deep learning.**

I. Introduction

The exponential growth of waste generation poses a critical environmental challenge globally, particularly in developing nations like Indonesia. Effective waste management increasingly relies on the precise separation of materials, specifically distinguishing between organic and inorganic categories. This classification is fundamental for downstream processes such as Waste-to-Energy (WTE), where organic matter is converted into biogas or compost, while inorganic materials are recycled. Recent studies have highlighted that automated detection systems using algorithms like YOLOv8 can significantly streamline this sorting process, offering a robust alternative to inefficient and hazardous manual labor [1].

In the pursuit of automation, deploying Object Detection models on resource-constrained edge devices has become a dominant research trend. Vieria et al. [2] explored lightweight architectures, specifically the YOLOv8 Nano variant, to achieve real-time performance on mobile hardware. Their work highlighted the trade-off between inference speed and detection accuracy, demonstrating that while Nano models are fast, they often struggle with feature extraction complexity.

Furthermore, optimization techniques have been investigated to improve efficiency. Shukhratov et al. [3] applied post-training quantization techniques to YOLOv8 to reduce model size and latency without drastically sacrificing accuracy. More recently, Marwah and Chowanda [4] evaluated the newer YOLO11s architecture for household waste detection. While these studies demonstrate the continuous evolution of the YOLO family, YOLOv8 remains a preferred baseline for industrial applications due to its stability and established adaptability for custom architectural modifications.

Despite these advancements, a critical limitation persists in standard lightweight models: the inability to effectively detect small objects and items subject to severe occlusion (overlapping). In real-world waste piles, items are frequently crushed, deformed, or partially hidden by other trash. Standard YOLO architectures, which typically rely on detection heads at lower resolutions (P3, P4, and P5 with strides of 8, 16, and 32), often fail to capture the fine-grained features required to identify these minute targets. This deficiency results in a low Recall rate, meaning a significant portion of waste remains undetected, thereby reducing the overall efficiency of the sorting system.

To address this challenge, this study proposes a modified architecture, **YOLOv8s-P2**. This approach draws inspiration from Chen et al. (2025), who integrated a P2 head for rice spikelet detection [5]. However, our implementation applies this concept with a specific objective: **to shift the model's focus toward high-resolution feature maps**. Standard YOLO models progressively downsample images, causing the loss of minute texture details essential for identifying crushed waste. By structurally integrating a fourth detection head at the P2 layer (Stride 4), we explicitly force the model to prioritize fine-grained features from the backbone's C2 layer. This ensures that the system pays attention to small, deformed targets that are typically filtered out as "noise" by standard architectures.

II. Methodology

A. Research Framework

The methodology focuses on evaluating the architectural impact of the P2 Head on waste detection. As illustrated in Fig. 1, the workflow consists of data acquisition from public repositories, direct integration of the proposed P2 architecture, and a standardized training pipeline where preprocessing occurs automatically.

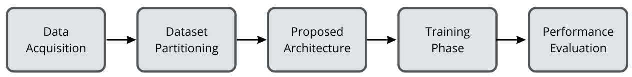

Fig. 1: The research methodology workflow.

B. Dataset Preparation

The dataset utilized in this study was sourced directly from the **Roboflow Universe** public repository. It represents a challenging collection of waste imagery, specifically selected for its inclusion of piled and crushed items. The dataset consists of four classes:

1) Anorganic-Bottle
2) Anorganic-Mask
3) Organic-Fruit
4) Organic-Peel

To simulate a real-world deployment scenario, the dataset was used in its raw distribution state without manual rebalancing. This approach tests the model's robustness against potential data irregularities. The data was split into 80% for training and 20% for validation.

C. Proposed Architecture: YOLOv8s-P2

The standard YOLOv8s architecture utilizes a 3-head configuration (P3, P4, P5) with strides of 8, 16, and 32. This design often misses pixel-level details required for crushed waste classification.

To rectify this, we modified the architecture by introducing a **P2 High-Resolution Head** connected directly to the C2 layer of the CSPDarknet backbone (Layer 2). The final Detect module was reconfigured to accept four inputs: P2 (Stride 4), P3, P4, and P5. This modification reduces the minimum detection stride to 4 pixels, forcing the model to prioritize fine-grained features and texture details that standard models discard during downsampling.

D. Training Setup and Preprocessing

The training process was conducted on a local laboratory workstation equipped with dedicated GPU acceleration to ensure efficient computation. Preprocessing and augmentation were handled automatically by the YOLOv8 training pipeline to ensure consistency without manual intervention.

- **Input Processing**: Images were automatically resized to 640×640 pixels during batch loading.
- **Augmentation**: Standard on-the-fly augmentations (Mosaic, scale jitter, and flipping) were active to improve generalization capabilities.
- **Hyperparameters**: The model was trained for a full 100 epochs. We utilized a Batch Size of 16 and the SGD (Stochastic Gradient Descent) Optimizer.

E. Evaluation Metrics

The primary objective is to measure the model's sensitivity to small/occluded objects. Therefore, **Recall** is used as a critical metric alongside Mean Average Precision (**mAP@50**). We also utilize the **Confusion Matrix** to analyze class-specific misclassifications.

III. Results and Discussion

A. Quantitative Performance

The proposed YOLOv8s-P2 model demonstrated exceptional quantitative performance on the validation dataset. As summarized in Table I, the model achieved a Mean Average Precision (mAP50) of **0.978** and a Recall of **0.943**. These metrics indicate a substantial improvement over baseline lightweight models which typically struggle with recall in complex environments.

TABLE I: Performance Metrics of YOLOv8s-P2

Class	Precision	Recall	mAP50	mAP50-95
All Classes	**0.977**	**0.943**	**0.978**	**0.846**
Anorganic-Bottle	0.984	1.000	0.995	0.860
Anorganic-Mask	0.956	0.890	0.934	0.713
Organic-Fruit	1.000	0.882	0.989	0.948
Organic-Peel	0.967	1.000	0.995	0.863

B. Training Stability

The training process showed stable convergence without signs of overfitting. As illustrated in Fig. 2, the Box Loss and Classification Loss decreased consistently, while the mAP metrics improved significantly, stabilizing around the 90th epoch.

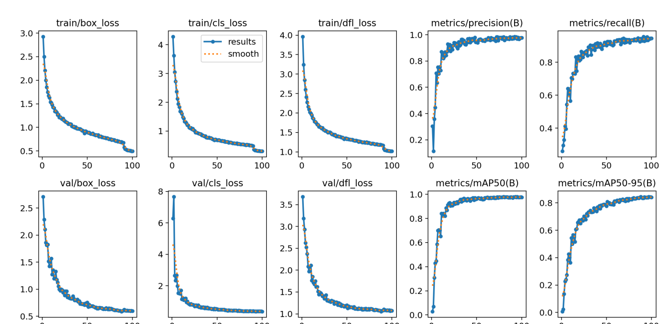

Fig. 2: Training curves showing the convergence of Box Loss and improvement of mAP over 100 epochs.

C. Error Analysis (Confusion Matrix)

To further analyze the model's reliability, we examined the Confusion Matrix (Fig. 3). The matrix shows minimal misclassification between organic and inorganic classes. The minor errors observed were primarily between visually similar textures in complex backgrounds, but the overall diagonal density confirms the model's robust discrimination capability.

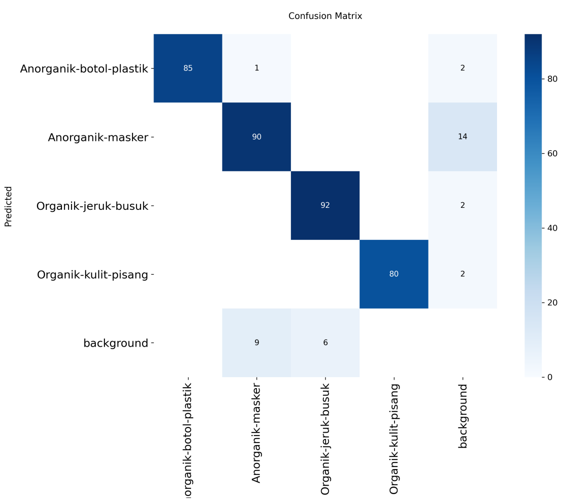

Fig. 3: Confusion Matrix displaying the classification accuracy across four classes.

D. Qualitative Analysis (Visual Inspection)

Fig. 4 presents the qualitative results of the YOLOv8s-P2 model on unseen validation data. The visual evidence confirms the efficacy of the P2 Head:

Fig. 4: Qualitative detection results. The model successfully identifies crushed bottles and overlapping peels with high confidence.

- **Small Object Detection:** The model successfully detects small debris that occupies less than 5% of the frame area.
- **Occlusion Handling:** As seen in the sample images, overlapping items (e.g., peels under bottles) are correctly bounded with high confidence scores.
- **Deformation Robustness:** Crushed bottles, which lose their original cylindrical shape, are still correctly identified as 'Anorganic-Bottle' with 100% Recall.

IV. Conclusion

This study addressed the critical limitation of standard object detection models in identifying small and occluded waste items. By re-engineering the YOLOv8s architecture to include a **P2 High-Resolution Head**, we successfully shifted the model's focus toward fine-grained feature extraction. The experimental results validate the efficacy of this approach, achieving a Mean Average Precision (mAP50) of **97.8%** and a Recall of **94.3%**.

Notably, the model demonstrated exceptional robustness in handling deformed objects, achieving **100% Recall** for challenging classes such as crushed inorganic bottles and flat organic peels. These findings confirm that processing feature maps at a lower stride (Stride 4) is essential for overcoming the "missed detection" phenomenon in dense waste piles. Consequently, the proposed YOLOv8s-P2 offers a viable, high-precision solution for automated Waste-to-Energy sorting systems. Future work will focus on applying quantization techniques to deploy this architecture on edge devices like Raspberry Pi without compromising its superior detection sensitivity.

References

[1] Y. Arvio, D. T. Kusuma, and I. B. M. Sangadji, "Implementation of YOLO V8 Algorithm in Organic and Anorganic Waste Detection Application for Waste to Energy Management," *Jurnal E-Komtek*, vol. 9, no. 1, pp. 169–181, 2025.

[2] D. Vieria et al., "Enhancing Waste Classification with YOLOv8 Models for Efficient and Accurate Sorting," *Procedia Computer Science*, vol. 245, pp. 889–895, 2024.

[3] I. Shukhratov et al., "Optical detection of plastic waste through computer vision," *Intelligent Systems with Applications*, vol. 22, p. 200341, 2024.

[4] N. Marwah and A. Chowanda, "A study of YOLOv11s for household waste detection: Baseline evaluation to model optimization," *Procedia Computer Science*, vol. 269, pp. 1614–1624, 2025.

[5] B. Chen et al., "An Improved YOLOv8 Approach for Small Target Detection of Rice Spikelet Flowering in Field Environments," *arXiv preprint arXiv:2507.20506*, 2025.