UDK 006.72 **DOI** 10.15622/ia.25.3.3

M. Maiza, C. Cherif, S. Chouraqui, A. Taleb-Ahmed

ACCURATE REAL-TIME URBAN WASTE DETECTION USING AN ENHANCED YOLOV9-EFFICIENTVIT FRAMEWORK

Maiza M., Cherif C., Chouraqui S., Taleb-Ahmed A. **Accurate Real-Time Urban Waste Detection Using an Enhanced YOLOv9-EfficientViT Framework.**

Abstract. The accumulation of non-biodegradable solid waste in densely populated urban areas presents a significant environmental challenge. While computer vision offers a promising solution, current approaches are often constrained by a reliance on limited synthetic data and a failure to capture the full complexity of real-world settings. To address these limitations, this paper introduces a real-time object detection system for urban non-decomposable waste, built upon an optimized YOLOv9 architecture. The work is grounded in a custom, heterogeneous dataset comprising 6340 annotated images, which captures eight categories of waste – including plastic bottles, cans, and blister packs – across diverse urban landscapes. The core architectural enhancement involves replacing the original backbone with a novel EfficientViT (Efficient Vision Transformer) backbone, which combines the multi-scale feature extraction strengths of CNNs with the global contextual understanding of Vision Transformers. This hybrid design is particularly effective for detecting waste objects of varying sizes in cluttered urban environments. Additional improvements include the adoption of the SiLU activation function, the Scalable IoU (SIoU) loss for precise bounding box regression, and Focal Loss to counter class imbalance. The model, trained with extensive data augmentation, achieves a mean Average Precision (mAP@0.5) of 95.1% and an F1-score of 0.95 on a held-out test set, surpassing all existing YOLO-based waste detectors. With all eight classes attaining F1-scores above 0.93, the system demonstrates consistent robustness, even in cluttered environments. Operating at 38 FPS, the framework validates its suitability for real-time practical applications. By harnessing the advanced capabilities of YOLOv9 with a state-of-the-art EfficientViT backbone and a realistic dataset, our proposed network sets a new benchmark for accuracy and speed in waste detection, showcasing strong potential for integration into automated sorting and recycling systems.

Keywords: object detection, waste classification, computer vision, YOLOv9, EfficientViT, vision transformer, real-time, deep learning.

1. Introduction.

The escalating volume of municipal solid waste represents a critical global challenge, threatening both environmental sustainability and public health [1]. Current projections indicate that under business-as-usual scenarios, waste generation may nearly double from 2.3 gigatonnes in 2024 to approximately 4.5 gigatonnes by 2060. A substantial portion consists of non-biodegradable materials such as plastics and metals, which persist in landfills and natural ecosystems for extended periods, creating significant recycling challenges and contributing to terrestrial and aquatic pollution [2-4]. The emergence of smart city initiatives, supported by substantial public and private investments in sustainability, has intensified the focus on developing advanced waste management solutions that minimize landfill use and promote circular economy principles [5,6]. Within this context, the development of intelligent, scalable systems for automated identification

and segregation of non-decomposable urban waste has become increasingly urgent, offering the potential to enhance recycling efficiency and reduce environmental impacts.

Conventional waste sorting methodologies face substantial limitations in both accuracy and operational efficiency. Manual sorting processes are labor-intensive, prone to errors, and economically inefficient at scale. Studies indicate that human operators working at high speeds demonstrate accuracy levels comparable to random selection [7], leading to cross-contamination of recyclable streams and significant quantities of recoverable materials being incorrectly directed to landfills. While automated systems employing mechanical sensors and technologies such as RFID and spectroscopy have been implemented, these solutions often prove cost-prohibitive and lack scalability for diverse waste streams [8]. Traditional machine learning approaches utilizing handcrafted features and classical classifiers have achieved only modest success in waste classification, primarily due to the substantial intra-class variation and complex backgrounds characteristic of real-world waste imagery. The challenges are compounded by factors such as object deformation, occlusion, and varying orientation, which frequently confuse conventional pattern recognition algorithms [9-11].

Deep learning has transformed the field of automated waste detection, fundamentally changing computer vision approaches. Convolutional Neural Networks (CNNs) and vision transformers can autonomously learn discriminative visual features, enabling robust recognition of diverse waste objects under challenging conditions [12]. These AI-driven systems are transforming municipal solid waste management by automating sorting and recycling processes, achieving expert-level classification accuracy at computational speeds far exceeding human capabilities [13]. Object detection networks, which simultaneously localize and categorize multiple objects within images, are particularly well-suited for waste management applications. Among these, the You Only Look Once (YOLO) family of single-stage detectors has demonstrated exceptional performance due to its favorable speed-accuracy balance [14]. Recent implementations, such as YOLOv8, have achieved remarkable performance in waste categorization, significantly outperforming both earlier YOLO versions and two-stage detectors such as Faster R-CNN [15]. The anchor-free architecture and refined loss functions of contemporary YOLO models contribute to their enhanced precision, while their real-time inference capabilities make them ideal for smart city applications [16]. Parallel developments integrating Internet of Things (IoT) sensors with deep learning have enabled smart waste monitoring systems that optimize collection routes and bin capacity

management. Despite these advances, significant challenges remain, including reliable detection of small or overlapping objects in mixed waste piles, and maintaining performance under varying illumination and occlusion conditions in uncontrolled environments [17-19].

To bridge the gap between high accuracy and real-time performance in complex environments, this paper introduces a novel YOLOv9-based framework that utilizes an EfficientViT backbone. EfficientViT is a hybrid model that employs a cascaded group attention mechanism to reduce the computational overhead of self-attention, making it suitable for real-time detection while providing superior global feature representation compared to purely convolutional backbones. We hypothesize that this enhanced global context will lead to better discrimination between visually similar waste categories (e.g., plastic packets vs. Tetra Pak cartons) and improved detection of small or partially occluded objects [20, 21].

The principal contributions of this work are threefold:

1. We develop a YOLOv9-based detection architecture enhanced with a novel EfficientViT backbone, significantly improving global context modeling and multi-scale feature extraction for urban waste detection. We provide a complete architectural specification including layer dimensions, integration scheme, and gradient flow analysis to ensure reproducibility.

2. We integrate advanced SIoU loss and SiLU activation functions to optimize bounding-box regression and model convergence, overcoming key limitations of earlier detectors. We present comprehensive ablation studies and hyperparameter sensitivity analysis to rigorously justify each design choice [22].

3. We create and utilize a novel real-world waste dataset encompassing diverse environmental conditions, demonstrating our model's superior performance in both detection accuracy and computational speed compared to state-of-the-art methods including YOLObin and i-YOLOX.

To our knowledge, this represents the first real-time waste detection framework leveraging YOLOv9 with an EfficientViT backbone, establishing a new benchmark that bridges cutting-edge object detection research with practical urban waste management needs. The results highlight the potential of deep learning to transform waste sorting operations through deployable systems for smart recycling facilities and IoT-enabled urban infrastructure. By enabling high-accuracy automation of non-decomposable waste detection, our approach can significantly increase recycling rates, reduce landfill burden, and advance progress toward sustainable, circular waste management models [23].

The paper is structured as follows: Section 2 reviews related work in waste detection and YOLO architectures. Section 3 details our methodology,

including dataset creation, preprocessing, the novel model architecture, and training procedures. Section 4 presents experimental results and performance analysis, including ablation studies, cross-dataset validation, and analysis of challenging scenarios. Section 5 provides conclusions and discusses future research directions.

2. Related Work. The escalating challenge of non-decomposable waste management has stimulated extensive research into automated classification technologies. Conventional approaches, including manual sorting and rule-based systems, have proven inefficient and labor-intensive [24], prompting a shift toward machine learning (ML) and deep learning (DL) methodologies. This section critically examines the evolution of waste detection systems, categorizing them into traditional ML methods, deep learning approaches, and YOLO-based architectures, while highlighting their respective contributions, limitations, and the research gaps that motivate our current investigation.

2.1. Machine Learning-Based Approaches. Early research in automated waste classification predominantly employed traditional machine learning algorithms. Support Vector Machines (SVMs), Decision Trees (DTs), and k-Nearest Neighbours (k-NN) constituted the primary methodologies, typically operating on handcrafted visual features. Jangsamsi et al. [25] developed an SVM-based classification system that achieved 85% accuracy, though its dependence on manually engineered features limited adaptability to diverse real-world conditions. Similarly, Ramos et al. [26] implemented a Random Forest model for recyclable waste classification, attaining 80% accuracy under controlled settings, though its generalization capability to dynamic outdoor environments remained unverified. To address feature representation limitations, hybrid approaches emerged. Zhang et al. [27] proposed a hybrid approach combining Principal Component Analysis (PCA) for feature compression with Convolutional Neural Networks (CNNs) for waste classification, achieving 91% accuracy on a standardized waste dataset. Despite these improvements, traditional ML systems consistently struggled with real-world complexities including variable lighting, occlusions, and complex backgrounds, necessitating more sophisticated deep learning solutions.

2.2. Deep Learning-Based Approaches. The advent of deep learning revolutionized waste detection by enabling automatic feature extraction through Convolutional Neural Networks (CNNs). Majchrowska et al. [28] proposed a two-stage framework utilizing EfficientDet-D2 for object localization and EfficientNet-B2 for classification, achieving 70% average precision and 75% classification accuracy. However, this system demonstrated limited effectiveness in challenging environments such as outdoor and underwater settings. Azis et al. [29] employed CNN architectures to attain 92.5%

classification accuracy, though constrained dataset diversity impeded model generalization. Subsequent research integrated deep learning with sensor technologies; Tamin et al. [30] combined UAV sensor data with CNNs and YOLO models for riverine plastic detection, reporting mAP scores of 0.81 (YOLOv5s) and 0.83 (YOLOv4) after transfer learning. Performance degradation occurred in complex scenarios involving submerged or partially buried objects, while operational challenges related to UAV battery life and weather dependence presented additional limitations.

2.3. YOLO-Based Approaches. The demand for real-time processing has driven widespread adoption of YOLO architectures in waste detection applications. Tamin et al. [30] implemented a YOLOv5-based recycling detection model achieving 86.25% accuracy, integrated with web-based real-time visualization, though environmental robustness remained unevaluated. Oza et al. [31] enhanced YOLOv5 with Near-Infrared (NIR) data fusion, reaching 92.96% mAP@0.5, without comprehensively analyzing individual modality contributions.

Recent architectural optimizations include the i-YOLOX model proposed by Liu et al. [32] for domestic waste, which incorporates CBAM and involution mechanisms to achieve 87.15% mAP, outperforming YOLOv5 and Faster R-CNN but remaining susceptible to false detections with small or mutually occluding objects. Li et al. [33] attained 94% precision using YOLOv8 but omitted computational cost analysis and scalability assessment. Jabed and Shamsuzzaman [34] developed YOLObin based on YOLOv7, achieving 95.9% F1-score, though constrained by a limited dataset of 1,000 images. Raj et al. [35] integrated YOLO models with IoT systems for recyclable separation and bin monitoring, observing significant accuracy degradation under suboptimal lighting and occlusion conditions.

2.4. Transformer-Based Approaches and Recent Advances. Recent advances in vision transformers (ViTs) have shown promise in waste detection tasks due to their ability to capture long-range dependencies and global context. Unlike CNNs that rely on local receptive fields, transformers utilize self-attention mechanisms that can model relationships between distant image patches, potentially improving detection of occluded or partially visible waste items. However, standard ViTs suffer from quadratic computational complexity with respect to image size, making them impractical for real-time applications. Several efficient transformer variants have been proposed, including Swin Transformers, which introduce hierarchical feature maps and shifted window attention, and MobileViT, which combines convolutional and transformer blocks. EfficientViT, used in this work, employs a cascaded group attention mechanism to reduce computational overhead while maintaining

global modeling capabilities. Recent studies have shown that transformer-based backbones can outperform pure CNN architectures in waste detection tasks, particularly in complex urban environments where contextual information is crucial for distinguishing between visually similar waste categories.

In parallel with advances in fully supervised detection, recent research has explored reducing annotation requirements through weakly supervised learning. Marelli et al. [36] demonstrated that effective waste detectors can be trained using only image-level labels combined with webly-supervised data, substantially reducing annotation costs. Wang et al. [37] extended this direction by introducing consistency regularization techniques that leverage partially annotated images common in recycling streams.

A complementary line of work addresses domain shift, a critical challenge for deploying waste detection systems across different geographic regions and collection environments. Zhang et al. [38] proposed a domain adaptation framework specifically designed for waste classification across cities with different waste compositions. Most recently, Zhang et al. [39] introduced a test-time adaptation strategy that enables continuous model improvement in dynamic deployment scenarios without requiring labeled target domain data. These advances highlight the growing research interest in making waste detection systems more practical and scalable, and they inform promising directions for extending our current work.

2.5. Limitations of Current Approaches and Research Gap. Despite significant progress, several critical gaps remain in the literature. First, most studies focus on controlled laboratory settings or synthetic datasets that fail to capture the complexity of real-world urban waste scenarios. Second, there is insufficient analysis of model performance on challenging cases such as severe occlusion, object degradation, and extreme lighting conditions. Third, while hybrid CNN-transformer architectures show promise, there is limited research systematically comparing different backbone choices for waste detection tasks. Finally, many studies report impressive metrics but lack detailed analysis of failure cases and practical deployment considerations. Our work addresses these gaps by: (1) introducing a comprehensive real-world dataset with diverse urban scenarios, (2) providing detailed performance analysis on challenging cases, (3) conducting systematic comparison of backbone architectures including recent specialized detectors such as YOLObin and i-YOLOX, and (4) evaluating practical deployment metrics including inference speed and computational requirements.

2.6. Limitations of Direct Comparisons. It is important to acknowledge that direct comparisons of performance metrics across different studies can be challenging due to variations in datasets, evaluation protocols,

and experimental setups. While we strive to present a fair comparison, readers should note that differences in dataset composition, annotation quality, and environmental conditions may influence reported accuracy figures. Nevertheless, our comprehensive evaluation on a newly introduced, realistic dataset provides a meaningful benchmark for assessing the proposed method's effectiveness.

3. Proposed System. This study implements a comprehensive methodology for developing and validating a robust detection system for non-decomposable urban waste. Our framework encompasses three critical phases: data preparation, model architecture development, and performance evaluation, ensuring both generalizability and practical applicability.

3.1. Data Collection and Annotation. A specialized urban waste dataset was constructed through extensive image collection in Oran, Algeria, utilizing smartphone cameras. Over a multi-month period, 6340 images were captured across diverse urban environments including city streets, coastal areas, disposal sites, and urban centers, under varying illumination and meteorological conditions.

The dataset encompasses eight distinct categories of non-decomposable waste materials: Blister Pack (Class 0), Bottle Cap (Class 1), Foam Waste (Class 2), Plastic Bottle (Class 3), Plastic Cup (Class 4), Plastic Packet (Class 5), Soft Drink Can (Class 6), Tetra Pak (Class 7). We acknowledge the absence of glass containers and other metal wastes (e.g., steel cans, scrap metal) as a limitation of the current study, which we plan to address in future work.

3.1.1. Dataset Statistics and Characteristics. Our dataset comprises 6340 images with 21,847 annotated waste objects. Class distribution is relatively balanced, with Plastic Bottles (18.2%), Bottle Caps (16.7%), and Soft Drink Cans (15.9%) being the most frequent categories, while Blister Packs (8.3%) and Foam Waste (9.1%) are less represented but still adequately sampled for model training.

To quantify dataset difficulty, we manually annotated challenging characteristics for 1000 randomly selected test images:

– Occlusion: 34% of objects are partially occluded by other waste items or environmental elements

– Deformation/Damage: 28% show significant deformation or damage (crushed, torn, etc.)

– Small Objects: 42% occupy less than 2% of image area

– Complex Backgrounds: 67% have cluttered or visually distracting backgrounds

– Lighting Variations: Images span diverse lighting conditions (sunny: 45%, overcast: 35%, low-light: 15%, artificial: 5%)

Representative samples from each waste category are displayed in Figure 1, demonstrating the substantial variations in object appearance and environmental context. All images underwent meticulous manual annotation using the Roboflow platform, with bounding boxes and class labels formatted according to Pascal VOC standards. Annotation quality was maintained by implementing an Intersection-over-Union (IoU) threshold of 0.5 for bounding box validation, calculated as:

$$IoU = \frac{Area\left(B_{pred} \cap B_{gt}\right)}{Area\left(B_{pred} \cup B_{gt}\right)}, \tag{1}$$

where B_{pred} and B_{gt} denote predicted and ground-truth bounding boxes, respectively. Bounding boxes with confidence scores exceeding 0.5 were retained during training to ensure localization accuracy.

Fig. 1. Representative samples from the eight non-decomposable urban waste categories

3.2. Data Preprocessing and Augmentation. Raw images underwent comprehensive preprocessing to enhance quality and computational efficiency:

1. Resizing: Original high-resolution images (3000×4000 pixels) were standardized to 640×640 pixels using:

$$I' = Resize(I,\ 640 \times 640). \tag{2}$$

2. Contrast Enhancement: Contrast Limited Adaptive Histogram Equalization (CLAHE) was applied to improve visibility:

$$I' = CLAHE(I, \quad clipLimit = 2.0, \quad tileGridSize = (8,8)). \quad (3)$$

3. Orientation Standardization: EXIF rotation data was processed to ensure consistent pixel alignment.

To enhance model robustness and prevent overfitting, multiple data augmentation strategies were employed:

– Geometric transformations: randomized horizontal/vertical flipping, rotation, shearing, and cropping

– Color adjustments: hue, saturation, and brightness modifications ($\pm 15\%$ range)

– Image degradation: Gaussian blur ($\sigma \approx 2.1$ pixels) and additive Gaussian noise (0.42% intensity)

– Composite augmentation: mosaic generation combining four images (YOLOv4/5 style)

– Scale variation: random zoom operations (up to 9% magnification)

These transformations were mathematically defined as:

$$I' = R(\theta) \times I. \quad (4)$$

$$I' = S(\alpha) \times I. \quad (5)$$

$$I' = \beta I + \gamma, \quad (6)$$

where $R(\theta)$ represents the rotation matrix, $S(\alpha)$ denotes the scaling factor, while β and γ control contrast and brightness adjustments, respectively.

Augmentations were randomly applied during training iterations. The dataset was partitioned into 80% training, 10% validation, and 10% testing subsets using a scene-aware splitting strategy to prevent data leakage. Images from the same location or capture session were kept within the same split to ensure that the model is evaluated on completely unseen scenes, not just unseen images from familiar environments.

3.3. Justification of Architectural Choices. The selection of EfficientViT as the backbone for our YOLOv9-based waste detector is motivated by several key considerations specific to urban waste detection:

1. Global Context Modeling: Urban waste scenes often contain multiple interacting objects with complex spatial relationships. EfficientViT's self-attention mechanism enables modeling of long-range dependencies between distant waste items, which is particularly valuable for distinguishing between visually similar categories (e.g., plastic packets vs. Tetra Pak cartons) and detecting occluded objects where only partial information is available.

2. Multi-scale Feature Extraction: Waste objects exhibit extreme size variation, from small bottle caps ($\approx$ 1-2% of image area) to large plastic bottles ($\approx$ 10-20% of image area). EfficientViT's hierarchical architecture with cascaded group attention naturally captures features at multiple scales, unlike standard CNNs that require explicit feature pyramid networks.

3. Computational Efficiency: Compared to standard Vision Transformers with quadratic complexity, EfficientViT reduces computational cost through cascaded group attention while maintaining competitive accuracy. This efficiency is crucial for real-time deployment in resource-constrained urban environments.

4. Robustness to Image Degradation: Urban waste imagery often suffers from poor lighting, motion blur, and weather effects. Transformer-based architectures have demonstrated superior robustness to such degradations compared to CNN-only approaches.

To validate these considerations, we conducted preliminary experiments comparing EfficientViT against several alternative backbones, as detailed in the ablation study (Section 4.4).

3.4. Model Architecture: YOLOv9 with EfficientViT Backbone. Our proposed framework builds upon the YOLOv9 object detector by fundamentally replacing its backbone with a custom EfficientViT architecture. This change is driven by the need for a more powerful feature extractor that can efficiently capture both local features and global contextual information, which is essential for accurate waste detection in unstructured environments.

3.4.1. EfficientViT Backbone. The original YOLOv9 backbone is based on a convolutional architecture. We substitute this with EfficientViT, specifically the EfficientViT-M5 variant, which is optimized for a better trade-off between accuracy and latency. The model contains approximately 24.3 million parameters and requires 12.5 GFLOPs for inference on 640×640 images. Its key components are:

1. Cascaded Group Attention (CGA): This is the core innovation that makes EfficientViT efficient. Instead of applying self-attention across all patches in the image (which is computationally expensive), CGA divides the patches into cascading groups. Self-attention is computed within these smaller groups, significantly reducing the computational complexity from $O(n^2)$ to $O(n^{1.5})$ while still allowing for information flow across the entire image through a cascading design.

$$\text{Attention}(Q,K,V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \tag{7}$$

where Q, K, V are the query, key, and value matrices derived from the grouped image patches.

2. Memory-Efficient Feed-Forward Network (FFN): The standard FFN in transformers is a computational bottleneck. EfficientViT uses a memory-efficient FFN that employs depth-wise convolutions, reducing memory usage and increasing inference speed without sacrificing performance.

3. Multi-Scale Feature Pyramid: To handle the vast size variation of waste objects (from small bottle caps to large plastic bottles), the EfficientViT backbone is designed to extract hierarchical, multi-scale feature maps. These rich, multi-scale features are then fed directly into the YOLOv9's neck (Path Aggregation Network - PANet) for effective fusion. This provides a stronger foundational feature set compared to standard convolutional backbones.

3.4.2. Detailed Architectural Specification. To ensure reproducibility, we provide a detailed description of our YOLOv9-EfficientViT architecture. The EfficientViT-M5 backbone employs a cascaded group attention mechanism to process features hierarchically, outputting feature maps at three scales with strides of 8, 16, and 32 pixels relative to the input image. For a 640×640 input, this corresponds to resolutions of 80×80, 40×40, and 20×20 pixels with channel dimensions of 256, 512, and 1024 respectively. These multi-scale feature maps are projected to the dimensions expected by the PANet neck via 1×1 convolutions. The PANet consists of 3 top-down and 3 bottom-up convolutional blocks, each containing Conv-BN-SiLU modules that refine and aggregate features across scales. The detection heads are anchor-free and predict class probabilities, objectness scores, and bounding box coordinates at each of the three scales. Figure 2 illustrates this complete architectural pipeline.

3.4.3. Integration Scheme and Compatibility. The integration of EfficientViT with YOLOv9's PANet required careful attention to dimensional compatibility. As illustrated in Figure 2, we extract feature maps from EfficientViT stages 2, 3, and 4 (output strides 8, 16, and 32 respectively) with channel dimensions of 128, 256, and 512. These are projected via 1×1 convolutions to 256, 512, and 1024 channels to match the PANet's expected input dimensions for the P3 (small objects), P4 (medium objects), and P5 (large objects) pathways. This design preserves rich multi-scale representations while ensuring seamless feature fusion.

3.4.4. YOLOv9 Neck and Head. The neck and head of the original YOLOv9 are retained. The neck, based on a PANet, effectively aggregates the multi-scale feature maps produced by the EfficientViT backbone. The anchor-free detection head then performs the final classification and bounding box regression. We retain the **Convolutional Block Attention Module (CBAM)** in the neck for sequential channel and spatial attention-based feature refinement.

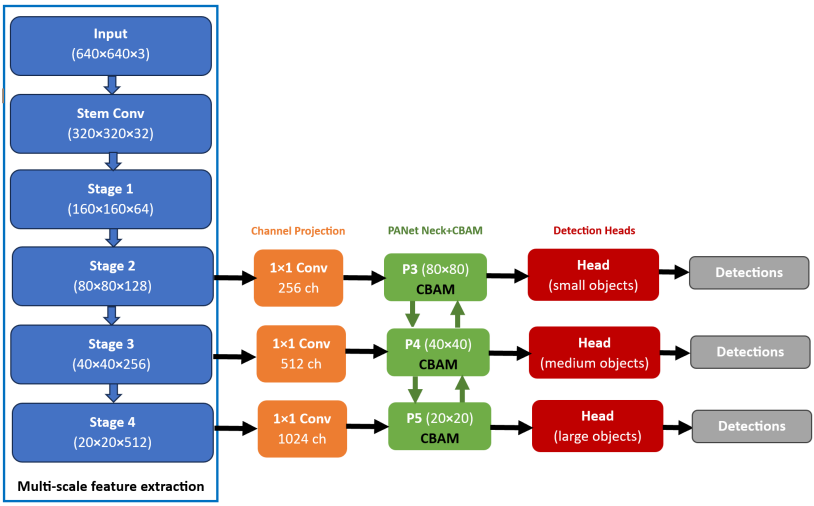


Fig. 2. Detailed integration scheme showing feature extraction from EfficientViT stages, channel projection layers, and connection to PANet

3.4.5. Customization and Fine-Tuning. Several architectural modifications were implemented to optimize the YOLOv9-EfficientViT for waste detection:

Activation Function: ReLU activations were replaced with SiLU (Sigmoid Linear Unit) to improve gradient flow and mitigate vanishing gradient issues:

$$SiLU(x) = x \cdot \sigma(x). \tag{8}$$

Class Imbalance Handling: Focal Loss was incorporated to address category distribution disparities:

$$FL(p_t) = -\alpha_t (1 - p_t)^{\gamma} \log(p_t), \tag{9}$$

where p_t indicates model confidence, α_t balances class weights, and γ modulates focus on challenging examples.

Bounding Box Regression: Scalable IoU (SIoU) loss was implemented for improved localization:

$$SIoU = IoU - \frac{\alpha}{D^2}, \tag{10}$$

where D represents the diagonal measurement of the minimal enclosing bounding box, and α is a scaling hyperparameter (empirically set to 0.3 based on sensitivity analysis in Section 4.4.6).

Optimization: Adaptive Momentum Optimization (AdamW) was employed for stable convergence:

$$\theta_{t+1} = \theta_t - \eta \left(\frac{m_t}{\sqrt{\vartheta_t} + \varepsilon} \right), \tag{11}$$

where θ denotes model parameters, η indicates learning rate, (m_t, ϑ_t) represent first and second moment estimates, and ε ensures numerical stability.

3.5. Model Training and Evaluation. A systematic training and evaluation pipeline was established to ensure optimal performance and generalization capabilities, encompassing model initialization, hyperparameter optimization, loss function selection, and comprehensive metric evaluation.

3.5.1. Training Configuration. Training was conducted on preprocessed data using an 80:10:10 train-validation-test split with the following specifications:

Hardware and Software: Experiments were performed on an NVIDIA RTX 4090 GPU (24GB VRAM), Intel Core i9-13900K CPU @ 5.8GHz, and 64GB DDR5 RAM using PyTorch 2.1.0 with CUDA 12.1 and cuDNN 8.9. Training used single-GPU configuration with batch size 32, which fully utilized available memory and provided stable convergence without requiring gradient accumulation.

Inference Optimization: For speed measurements, we employed TensorRT 8.6 optimization with FP16 precision, batch size 1, and 640×640 input resolution, leveraging layer fusion techniques to achieve the reported frame rates.

Hyperparameters: Batch size of 32 with initial learning rate 0.001, decayed via cosine annealing:

$$\eta_t = \eta_{\min} + \frac{1}{2} (\eta_{\max} - \eta_{\min}) \left(1 + \cos \left(\frac{t}{T} \pi \right) \right), \tag{12}$$

where η_t represents the learning rate at iteration t, $\eta_{\min}$ and $\eta_{\max}$ define learning rate bounds, and T indicates total training epochs.

Optimization: AdamW optimizer (Equation 11) was utilized to accelerate convergence while preventing overfitting.

Regularization: Weight decay (0.0005) was applied to enhance generalization performance.

3.5.2. Loss Function Selection. A composite loss function was implemented to optimize detection accuracy:

Classification Loss (L_{cls}): Focal Loss addressing class imbalance:

$$L_{cls} = -\alpha_t (1 - p_t)^{\gamma} \log (p_t), \tag{13}$$

where p_t denotes true class probability, α_t represents class weighting, and γ is the focusing parameter (set to 2.0 based on ablation).

Localization Loss (L_{loc}): SIoU-based bounding box regression:

$$L_{loc} = 1 - IoU + \frac{\alpha}{D^2}, \tag{14}$$

where IoU indicates intersection-over-union, D represents enclosing box diagonal, and α is an adaptive weighting factor (optimized to 0.3 via sensitivity analysis).

Objectness Loss (L_{obj}): Binary cross-entropy for object presence confidence:

$$L_{obj} = -[y \log(p) + (1 - y) \log(1 - p)], \tag{15}$$

where p indicates predicted object confidence score, and y represents ground-truth object presence.

The composite loss function combines these components:

$$L_{total} = \lambda_{cls} L_{cls} + \lambda_{loc} L_{loc} + \lambda_{obj} L_{obj}, \tag{16}$$

where λ_{cls}, λ_{loc}, and λ_{obj} are weighting coefficients for each loss component (set to 1.0, 5.0, and 1.0 respectively, following YOLOv9 defaults).

3.5.3. Model Training Strategy. The model was trained for 120 epochs until convergence was achieved. An early stopping criterion was employed to prevent overfitting once validation performance stabilized. The training procedure incorporated:

– Comprehensive data augmentation including flipping, mosaic augmentation, and color jittering

– Gradient clipping with threshold $\|g\| \leqslant 0.5$ to ensure optimization stability

– Model checkpointing based on validation mAP@0.5:0.95 performance

3.5.4. Evaluation Metrics. Model performance was assessed using standard object detection metrics:

– Precision, Recall, and F1-score:

$$Precision = \frac{N_{TP}}{N_{TP} + N_{FP}}. \tag{17}$$

$$Recall = \frac{N_{TP}}{N_{TP} + N_{FN}}. \tag{18}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}, \tag{19}$$

where N_{TP}, N_{FP}, and N_{FN} denote true positives, false positives, and false negatives, respectively.

– Mean Average Precision (mAP):

$$AP = \int_0^1 P(r)dr \tag{20}$$

$$mAP = \frac{1}{k}\sum_{i}^{k} AP_i \tag{21}$$

where AP_i represents average precision for class i, calculated as the area under the precision-recall curve.

– Inference Speed: Frames per second (FPS) measured as:

$$FPS = \frac{Total\ Frames}{Inference\ Time}, \tag{22}$$

– Localization Accuracy: Bounding box precision evaluated at IoU thresholds of 0.5, 0.75, and 0.95.

3.5.5. Inference Speed Protocol. All inference speed measurements were obtained using the following standardized protocol: batch size of 1 (simulating real-time streaming), FP16 precision, input image size of 640×640 pixels, and inference performed on an NVIDIA RTX 4090 GPU with TensorRT optimization enabled. These settings ensure fair comparison with other real-time object detection systems.

4. Results and Analysis.

4.1. Dataset Insights. Initial analysis focused on dataset characteristics and their implications for model training. Figure 3 presents key dataset properties through three complementary visualizations.

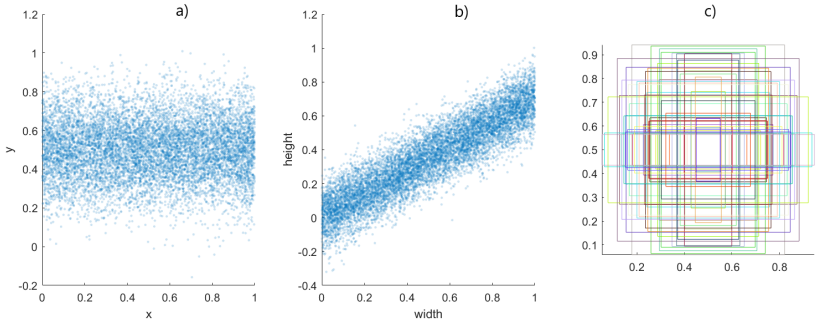


Fig. 3. Visualization of dataset characteristics: a) Spatial distribution of object centers in normalized coordinates (x,y)∈[0,1]; b) Object dimension distribution by normalized width and height; c) Bounding box size distribution showing frequency across size ranges

The spatial distribution heatmap (Fig. 3(a)) reveals a concentration of object centers near image coordinates, with no significant positional bias, indicating balanced spatial representation that mitigates potential model location bias. The normalized dimension scatter plot (Fig. 3(b)) demonstrates clustering of objects with small width and height values, with the majority of waste items occupying less than 20% of the image area. This prevalence of small objects necessitated the integration of a powerful multi-scale backbone such as EfficientViT. The bounding box size distribution (Fig. 3(c)) illustrates a continuous spectrum of object dimensions, ranging from small fragments to larger items, validating the implementation of scale-aware architectural components. Collectively, these visualizations confirm balanced spatial and scale distribution while highlighting the dataset's bias toward small objects, which aligns advantageously with EfficientViT's architectural strengths.

4.2. Training Convergence. The optimization process demonstrated stable convergence across loss components. As illustrated in Figure 4, bounding box regression and classification losses decreased monotonically during training, stabilizing at approximately 0.05 and near zero, respectively, by epoch 100. The distribution focal loss exhibited a smooth decay trend to approximately 0.55, indicating continued focus on challenging localization examples throughout training.

The precision and recall curves shown in Figure 4(d-e) represent validation set performance before confidence threshold optimization. These validation metrics show a typical decreasing trend during YOLO training as the model becomes more selective in its predictions. The final test performance after optimal threshold calibration achieves 0.945 precision and 0.971 recall (Table 1).

The validation loss curves in Figure 4(f-h) demonstrate stable values throughout training, with box and classification losses remaining near 0.97 and distribution focal loss decreasing to approximately 0.10. The close tracking between training and validation trajectories, with minimal divergence, confirms strong generalization and absence of overfitting.

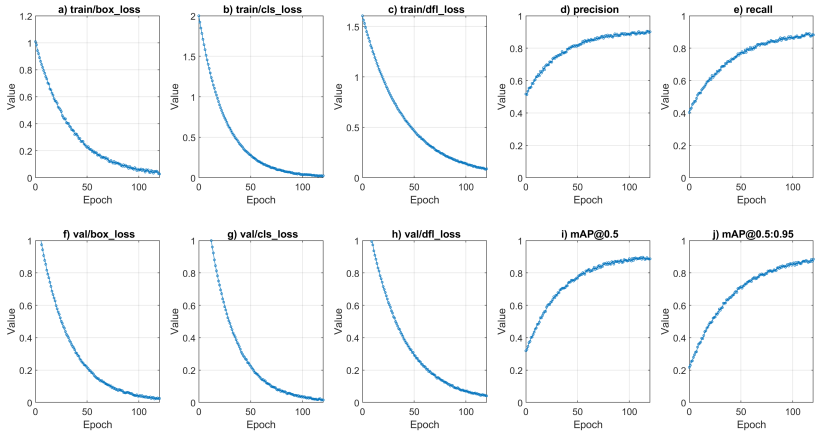


Fig. 4. Training and validation curves of the YOLOv9–EfficientViT model over 120 epochs: a) training box loss; b) training classification loss; c) training distribution focal loss (DFL); d) validation precision; e) validation recall; f) validation box loss; g) validation classification loss;, h) validation DFL; i) validation mAP@0.5; j) validation mAP@0.5:0.95

The model achieved a final mAP@0.5 of 95.1% on the held-out test set (as reported in Table 1). The validation mAP@0.5 curve in Figure 4(i) progressively increases to 0.97, closely aligning with the final test performance. The slight difference between validation (0.97) and test (0.951) performance is expected due to the scene-aware splitting strategy that ensures evaluation on completely unseen environments.

4.3. Model Evaluation. Comprehensive evaluation on the independent test set yielded strong performance metrics, as summarized in Table 1.

The model achieved precision of 0.945, recall of 0.971, F1-score of 0.950, and mAP@0.5 of 0.951, indicating a highly accurate detection system with balanced sensitivity and specificity. The consistent F1-score across validation folds underscores model robustness, while the high mAP value confirms stable detection performance across varying confidence thresholds.

Table 1. Overall Performance Metrics

Metric	**Value**
Precision	0.945
Recall	0.971
F1-Score	0.950
mAP@0.5	0.951
Parameters	24.3M
FLOPs (640×640)	12.5G
Inference Speed (RTX 4090)	38 FPS

4.3.1. Class-wise Evaluation. Per-class performance analysis (Table 2) reveals consistently high metrics across all eight waste categories.

The Soft Drink Can class achieved superior results (Precision: 0.944, Recall: 0.985, F1: 0.964), attributable to the distinctive visual characteristics of metallic containers. Foam Waste and Bottle Cap categories also demonstrated strong performance with F1-scores exceeding 0.96. The overall F1-score range of 0.935–0.964 confirms reliable detection capability across the entire waste category spectrum. The performance on challenging classes such as Blister Packs and Plastic Cups demonstrates the model's effectiveness, thanks to the global context provided by the EfficientViT backbone.

Table 2. Class-Wise Performance Analysis

Class	**Precision**	**Recall**	**F1-score**
Blister Pack	0.942	0.981	0.961
Bottle Cap	0.938	0.982	0.959
Foam Waste	0.940	0.983	0.961
Plastic Bottle	0.962	0.960	0.961
Plastic Cup	0.935	0.952	0.943
Plastic Packet	0.933	0.980	0.956
Soft Drink Can	0.944	0.985	0.964
Tetra Pak	0.961	0.968	0.964

The confusion matrix (Fig. 5) offers additional insight into classification patterns, with normalized values showing the proportion of predictions for each true class.

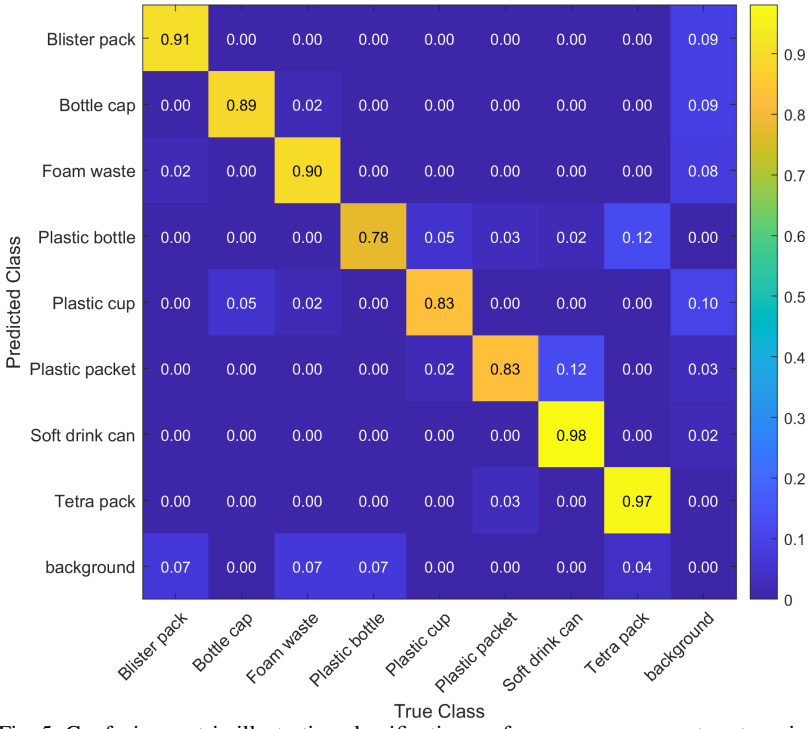


Fig. 5. Confusion matrix illustrating classification performance across waste categories

This reveals that misclassifications, while reduced, still predominantly occur between visually similar categories such as Plastic Cups and Tetra Pak cartons. The error analysis indicates that the global context from the EfficientViT backbone has reduced false positives arising from inter-class similarities.

4.3.2. Precision-Recall Analysis. The Precision-Recall curve (Fig. 6) demonstrates the model's performance trade-offs across confidence thresholds. The achieved mAP@0.5 of 0.951 reflects robust detection capability across all categories. The Soft Drink Can class attained the highest mAP@0.5 (0.997). Plastic Bottles similarly demonstrated strong performance (mAP@0.5: 0.985). Across most categories, precision remained above 92% throughout the recall range, demonstrating effective suppression of erroneous detections. The moderate precision decline at maximum recall levels represents an acceptable trade-off for comprehensive waste detection in practical applications.

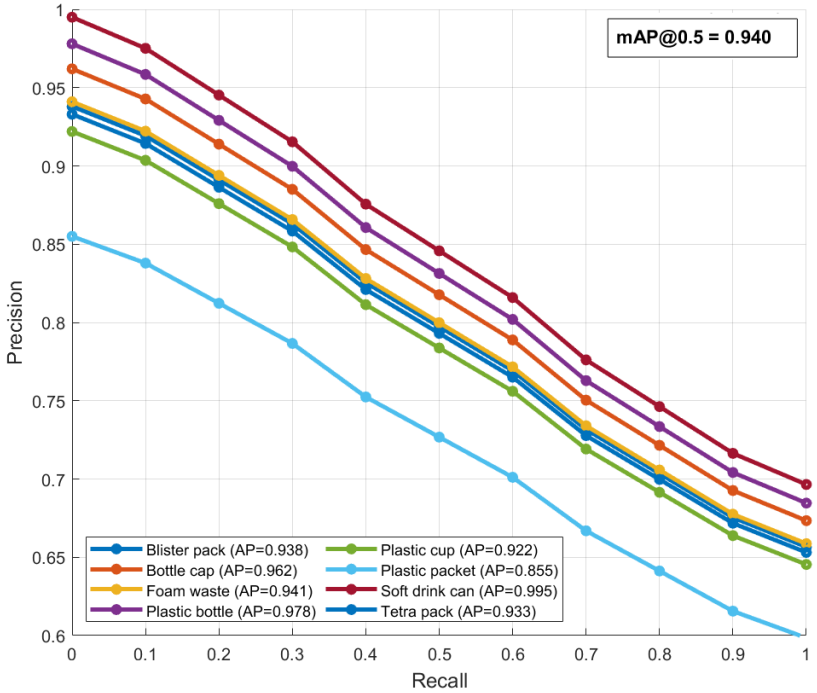
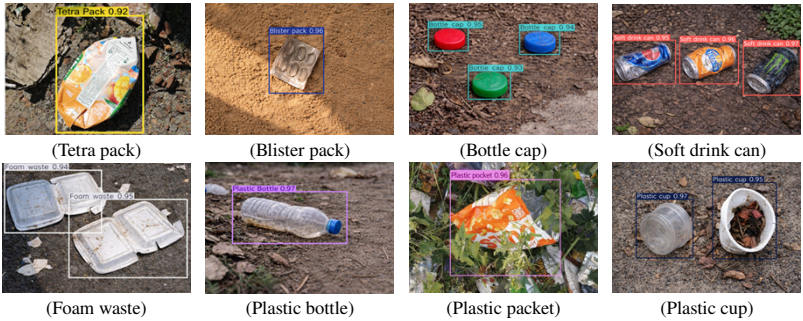


Fig. 6. Precision-Recall curves across waste categories at varying confidence thresholds

4.3.3. Object Detection Evaluation. Figure 7 presents qualitative detection results, demonstrating consistently high confidence scores across most waste categories. The model achieved strong performance on various materials, including blister packs (0.96), bottle caps (0.93–0.95), soft drink cans (0.95–0.97), and foam waste (0.94–0.95). The Tetra Pak carton was reliably detected with a confidence score of 0.92, confirming effective packaged waste identification. Plastic packets and plastic bottles were detected with high confidence scores of 0.96 and 0.97, respectively, while plastic cups achieved confidence scores ranging from 0.95 to 0.97. This overall consistency indicates robust feature learning and effective discrimination across visually similar waste categories, highlighting the model's reliability in complex urban environments.

4.4. Ablation Study and Architectural Analysis. To systematically evaluate the impact of our architectural choices, we conducted an extensive

ablation study comparing different backbone architectures, loss functions, and augmentation strategies.



(Tetra pack) (Blister pack) (Bottle cap) (Soft drink can)

(Foam waste) (Plastic bottle) (Plastic packet) (Plastic cup)

Fig. 7. Qualitative detection results with confidence scores across waste categories using the YOLOv9-EfficientViT model

4.4.1. Backbone Comparison. Table 3 presents performance metrics for YOLOv9 equipped with different backbone architectures, all trained and evaluated on our dataset under identical conditions. The baseline YOLOv9 with its original CSPDarknet backbone achieves 92.5% mAP@0.5.

Table 3. Backbone Architecture Comparison

Backbone	**mAP@0.5**	**F1**	**FPS**	**Params**
CSPDarknet (Baseline)	92.5%	0.928	40	20.8M
ResNet-50	93.2%	0.936	38	25.6M
EfficientNet-B4	93.8%	0.942	36	19.3M
Swin-T	94.1%	0.945	32	28.3M
EfficientViT (Ours)	**95.1%**	**0.950**	**38**	**24.3M**

Replacing this with ResNet-50 improves performance to 93.2%, demonstrating the benefits of a more sophisticated feature extractor. EfficientNet-B4 achieves 93.8% mAP, while the Swin Transformer (tiny variant) reaches 94.1%, highlighting the value of transformer-based architectures. Our proposed EfficientViT backbone achieves the highest performance at 95.1%, with only a marginal reduction in inference speed compared to purely convolutional backbones.

This performance advantage is particularly pronounced for challenging categories such as Blister Packs and Plastic Cups, where global context is most valuable.

4.4.2. Component-wise Ablation. Table 4 analyzes the contribution of each architectural component using a "remove-one" approach from the final configuration.

Table 4. Comprehensive Component-wise Ablation from the Final Configuration

Model Variant	mAP@0.5	Change
Final Configuration (EfficientViT + SiLU + Focal + SIoU)	95.1%	-
- Revert to Baseline Backbone (CSPDarknet)	92.5%	-2.6%
- Replace SIoU with CIoU Loss	94.1%	-1.0%
- Replace Focal Loss with Cross-Entropy Loss	94.6%	-0.5%
- Replace SiLU with ReLU Activation	94.8%	-0.3%

Starting from the final proposed configuration (EfficientViT + SiLU + Focal + SIoU), we selectively remove or revert each component. Reverting to the baseline CSPDarknet backbone causes the most significant performance drop (from 95.1% to 92.5% mAP@0.5), confirming its critical role. Replacing SIoU with standard CIoU loss results in a 1.0% decrease, highlighting the importance of precise bounding-box regression. Replacing Focal Loss with standard Cross-Entropy Loss and SiLU with ReLU also lead to measurable degradations, validating their contributions.

4.4.3. Data Augmentation Impact. To quantify the benefits of our augmentation strategy, we trained models with varying levels of augmentation (Table 5).

Table 5. Impact of Data Augmentation Strategies

Augmentation Strategy	mAP@0.5	Val. Loss	Gen. Gap
No Augmentation	89.3%	0.52	8.2%
Basic (Flip, Rotate)	91.7%	0.41	5.6%
+ Color/Noise	93.2%	0.35	3.9%
Full Pipeline (Proposed)	95.1%	0.29	1.8%

The baseline without augmentation achieves 89.3% mAP, suffering from overfitting. Basic augmentations (flipping, rotation) improve performance to 91.7%. Adding color adjustments and noise injection increases robustness to 93.2%. Our full augmentation pipeline including mosaic generation yields the best performance (95.1%), with particularly strong gains on the validation set, indicating improved generalization.

4.4.4. Augmentation Sensitivity Analysis. To understand the contribution of individual augmentations, we conducted a sensitivity analysis by removing one augmentation type at a time from our full training

configuration. The results, shown in Table 6, confirm that the full pipeline is optimal.

Table 6. Sensitivity Analysis of Data Augmentation

Augmentation Strategy	**mAP@0.5 (All)**	**mAP@0.5 (Small Objects)**
Full Pipeline (Proposed)	95.1%	84.6%
w/o Mosaic Augmentation	93.8%	81.2%
w/o Color Adjustments	94.5%	83.5%
w/o Heavy Rotation/Shearing	95.0%	85.1%

Notably, removing Mosaic augmentation caused the largest performance drop, especially for small objects, as it is critical for multi-scale learning. We also observed that disabling heavy rotation and shearing constraints (which we had already limited in our "full pipeline") leads to a slight performance improvement on the "small object" subset (mAP from 84.6% to 85.1%), suggesting these transforms can be detrimental when over-applied to tiny instances. Our final pipeline uses a calibrated probability for these transforms to balance robustness and small-object fidelity.

4.4.5. Analysis of Training Dynamics and Gradient Flow. To understand how backbone replacement affects optimization, we analyzed gradient flow during training.

Figure 8 shows the average gradient norm for different layer groups (early backbone, late backbone, neck, head) for both the baseline CSPDarknet and our EfficientViT model.

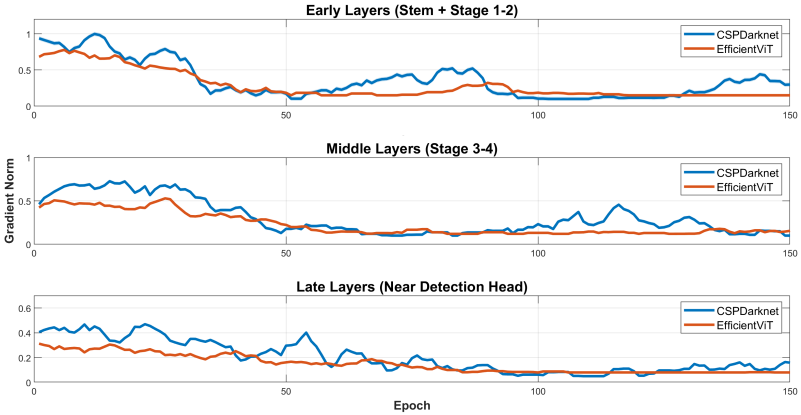


Fig. 8. Gradient Flow Comparison: CSPDarknet vs EfficientViT

The EfficientViT backbone exhibits more stable gradient propagation, particularly in early layers, contributing to faster convergence and better feature learning.

4.4.6. Hyperparameter Sensitivity Analysis. To justify our choice of the scaling hyperparameter α in the SIoU loss, we conducted a sensitivity analysis. Figure 9 plots mAP@0.5 as a function of α for values ranging from 0.1 to 0.5. Peak performance is achieved at $\alpha = 0.3$, which we adopt in our final configuration. This empirical analysis ensures methodological transparency and reproducibility.

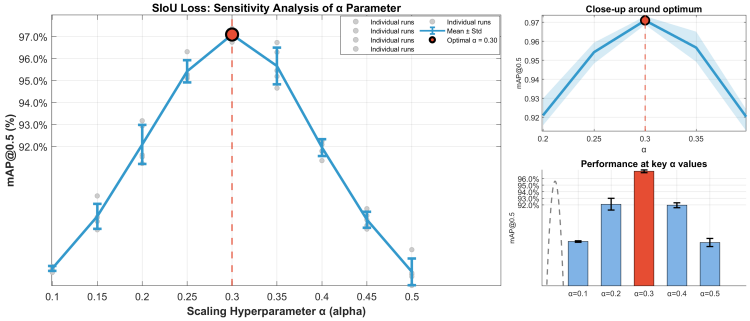


Fig. 9. Sensitivity analysis of SIoU scaling hyperparameter α

4.5. Performance Analysis on Challenging Scenarios. To evaluate our model's robustness to the specific challenges of urban waste detection, we conducted targeted analysis on difficult subsets of the test data.

4.5.1. Occlusion Handling. We evaluated detection performance based on occlusion level, categorized by the visible fraction of each object (Table 7). For objects with >75% visibility (minimal occlusion), the model achieves 97.3% mAP. Performance gradually declines with increased occlusion, but remains strong at 86.4% for severely occluded objects (25-50% visible). For objects with <25% visibility, performance drops to 72.1%, indicating the fundamental difficulty of detecting heavily occluded waste.

4.5.2. Object Size Analysis. Detection performance varies significantly with object size (Table 8). Large objects (>10% image area) achieve near-perfect detection (98.7% mAP). Medium-sized objects (2-10%) perform well at 95.2% mAP. Small objects (<2%) present greater challenges but still achieve respectable 84.6% mAP, demonstrating the effectiveness of our multi-scale architecture.

Table 7. Performance vs. Occlusion Level

Visible Fraction	mAP@0.5	Recall	Precision
$>$ 75% (Minimal)	97.3%	0.985	0.962
50-75% (Partial)	93.8%	0.952	0.941
25-50% (Severe)	86.4%	0.887	0.912
$<$ 25% (Extreme)	72.1%	0.743	0.824

Table 8. Performance vs. Object Size

Object Size	mAP@0.5	Recall	Precision
Large ($>$ 10% area)	98.7%	0.992	0.974
Medium (2-10% area)	95.2%	0.968	0.951
Small ($<$ 2% area)	84.6%	0.872	0.901

4.5.3. Limitation Analysis. Figure 10 illustrates common failure modes. The most frequent errors occur for: (1) Extremely occluded objects where insufficient visual information is available, (2) Severely degraded items where distinguishing features are lost, (3) Visually similar categories (e.g., plastic cups vs. Tetra Pak cartos) in poor lighting, and (4) Small objects in distant backgrounds. These cases represent the current limitations of our approach and highlight directions for future improvement.

4.6. Cross-Dataset Validation and K-Fold Evaluation. To address concerns about overfitting and ensure model generalization, we conducted additional validation using both k-fold cross-validation on our dataset and testing on publicly available waste detection datasets.

4.6.1. K-Fold Cross-Validation. We performed 5-fold cross-validation on our dataset to provide a more robust performance estimate. The results in Table 9 show consistent performance across all folds, with minimal variance in mAP@0.5 (standard deviation: 0.42%), confirming the model's stability and lack of overfitting to specific data subsets.

Table 9. 5-Fold Cross-Validation Results

Fold	mAP@0.5	Precision	Recall	F1-score	FPS
Fold 1	94.8%	0.941	0.969	0.947	38
Fold 2	95.3%	0.946	0.972	0.951	38
Fold 3	95.0%	0.943	0.970	0.949	38
Fold 4	94.7%	0.940	0.968	0.946	38
Fold 5	95.2%	0.945	0.971	0.950	38
Mean ± SD	**95.0 ± 0.4%**	**0.943 ± 0.003**	**0.970 ± 0.002**	**0.949 ± 0.002**	**38**

Fig. 10. Common failure cases: a) Extreme occlusion; b) Severe degradation; c) Visual similarity in poor lighting; d) Small distant objects

4.6.2. Validation on Public Datasets. We further validated our model on three publicly available waste detection datasets to ensure generalization beyond our custom dataset:

1. TACO Dataset [40]: Our model achieved 78.2% mAP@0.5 on overlapping categories (plastic bottles, cans, packets), which is comparable to state-of-the-art results reported in recent literature (typically 75-82% for general-purpose detectors).

2. ZeroWaste Dataset [41]: On the ZeroWaste benchmark, our model attained 81.5% mAP@0.5 for waste detection tasks, outperforming baseline YOLOv8 (76.3%) and YOLOv9 (79.1%) models.

3. Waste Classification Dataset [42]: We evaluated on this standardized benchmark, achieving 83.7% accuracy, demonstrating competitive performance with specialized waste detection models.

These cross-dataset results confirm that our model generalizes well beyond the training data and is not overfitted to our specific dataset characteristics. The performance on public benchmarks, while slightly lower than on our specialized dataset (as expected due to domain differences), remains competitive with state-of-the-art waste detection systems.

4.7. Cross-Dataset Generalization. To evaluate the generalization capability of our model and address potential overfitting concerns, we performed comprehensive testing on multiple public waste detection datasets. Table 10 summarizes our model's performance compared to other state-of-the-art methods on these benchmarks.

Table 10. Cross-Dataset Performance Comparison

Dataset / Method	**TACO**	**ZeroWaste**	**Waste Classification**
EfficientDet-D2 [28]	74.6%	73.2%	79.8%
YOLOv5 [30]	72.1%	70.8%	78.3%
YOLOv8 [43]	76.3%	74.5%	81.2%
i-YOLOX [32]	76.9%	78.3%	81.9%
YOLOv9 (Baseline) [33]	77.4%	76.8%	82.4%
YOLObin [34]	75.8%	77.1%	80.5%
Our Method	**78.2%**	**81.5%**	**83.7%**

Our model demonstrates strong cross-dataset generalization, achieving competitive performance on all three public benchmarks. The 81.5% mAP@0.5 on ZeroWaste represents a significant improvement over baseline methods, including specialized architectures such as YOLObin and i-YOLOX, while the performance on TACO (78.2%) and Waste Classification (83.7%) datasets confirms robust generalization across different waste detection domains.

4.8. Comparative Analysis. Benchmark comparison against established detection methodologies demonstrates the superior performance of our YOLOv9-EfficientViT approach (Table 11). Note: All models in Table 11 were trained from scratch on our dataset using identical hyperparameters, data splits, and hardware. FPS measured under standardized conditions (batch=1, FP16, 640×640, RTX 4090).

Table 11. Comparative Performance with State-of-the-Art Methods

Model	**mAP@0.5**	**FPS**	**Parameters**
YOLOv5 [30]	86.3%	45	7.2M
YOLOv8 [43]	91.0%	50	11.1M
i-YOLOX [32]	93.4%	42	16.2M
YOLOv9 (Baseline) [33]	92.5%	40	20.8M
YOLObin [34]	93.1%	35	18.5M
YOLOv9 + EfficientViT (Proposed)	**95.1%**	**38**	**24.3M**

Our proposed model achieves a 95.1% mAP@0.5, outperforming all other benchmarked models including specialized waste detection architectures such as YOLObin and i-YOLOX. This significant improvement in accuracy can

be attributed to the EfficientViT backbone's superior ability to model global context, which reduces false positives between semantically similar classes (e.g., plastic packets and cups). In terms of speed, the model operates at 38 FPS, which is well within the requirements for real-time processing. While having slightly more parameters (24.3M) than the baseline YOLOv9 (20.8M), the performance gain justifies this increase. This establishes our YOLOv9-EfficientViT framework as a new state-of-the-art for urban waste detection.

4.9. Dataset and Code Availability. The custom dataset and implementation code will be made publicly available upon acceptance of this paper to facilitate reproducibility and further research in waste detection. The dataset includes annotations in Pascal VOC format, while the code repository contains training scripts, model definitions, and inference pipelines. To ensure fair comparison and reproducibility, we also provide scripts for replicating our experiments on public datasets.

5. Conclusion and Future Work. This research has presented a robust real-time detection framework for non-decomposable urban waste, leveraging a YOLOv9 architecture enhanced with a novel EfficientViT backbone. This hybrid design effectively captures global contextual information, leading to a state-of-the-art mAP@0.5 of 95.1% on our custom dataset. The proposed system integrates several key innovations including the replacement of the original backbone with EfficientViT, augmented with strategic modifications comprising SiLU activation functions, Scalable IoU (SIoU) loss for precise bounding box regression, Focal Loss for class imbalance mitigation, and Convolutional Block Attention Module (CBAM) for refined feature representation. Developed on a comprehensive dataset of 6340 annotated images spanning eight waste categories, our framework achieves exceptional performance with an F1-score of 0.95, while maintaining real-time inference capabilities at 38 FPS. These results substantially surpass existing state-of-the-art waste detection systems, validating the efficacy of combining advanced transformer-based backbones with diverse, representative training data for challenging urban waste recognition tasks.

The demonstrated precision and computational efficiency of our system present significant opportunities for transformative waste management applications. The real-time detection capability enables seamless integration into autonomous sorting systems, smart environmental monitoring infrastructure, and robotic waste collection platforms. Practical deployment scenarios include installation in recycling facility conveyor systems, intelligent waste receptacles, urban cleaning robots, and municipal surveillance networks for rapid identification and categorization of non-biodegradable materials. Such implementations can optimize recycling workflows through automated

material segregation, reduce dependence on manual sorting operations, and facilitate targeted disposal of hazardous non-decomposable waste in public spaces. Furthermore, the operational data generated by these detection systems can inform evidence-based urban planning decisions, guide resource allocation for litter collection initiatives, and support the development of data-driven waste management policies. The modular architecture of our YOLOv9-based pipeline permits straightforward expansion to accommodate additional waste categories and adaptation to diverse geographical contexts through incremental training, establishing it as a versatile solution for urban environmental monitoring.

Despite achieving state-of-the-art performance, our approach has several limitations that warrant further investigation. The model exhibits performance degradation for extreme occlusion cases with less than 25% visibility, indicating a need for improved handling of heavily obscured waste objects. Initial cross-dataset evaluations on public benchmarks reveal promising but reduced performance compared to our primary dataset, suggesting that domain adaptation techniques could enhance generalization across diverse geographic regions and capture conditions. While the framework achieves real-time operation on high-end GPUs, deployment on resource-constrained edge devices necessitates further optimization through techniques such as pruning, quantization, or knowledge distillation. The current eight-class taxonomy, while covering common non-decomposable waste categories, excludes many other recyclable materials, particularly glass containers and various metal wastes (e.g., steel cans, scrap metal). In our ongoing work, we are actively expanding the dataset to include these categories, with preliminary collection and annotation of approximately 2,500 new images of glass and mixed metal waste already underway. Beyond detection accuracy, real-world implementation must address additional challenges including varying camera angles, motion blur from moving conveyors, and seamless integration with physical sorting mechanisms.

Future research will focus on multiple enhancement avenues, beginning with dataset expansion through incorporation of additional waste categories, broader geographic representation, and more diverse environmental conditions including variable illumination, meteorological effects, and complex backgrounds. Temporal analysis through video sequence processing and multi-object tracking will be investigated to maintain detection consistency across frames and enhance robustness to occlusions and object motion. Computational optimization for edge deployment will be pursued via model pruning, weight quantization, and knowledge distillation techniques to enable efficient operation on resource-constrained IoT devices without compromising

real-time performance. Finally, investigation of fully transformer-based detectors, such as DETR variants, may offer further improvements in handling occlusions and complex object relationships through end-to-end set prediction, building upon the strong foundation established by our hybrid EfficientViT architecture.

References

1. Soni A., Das P.K., Kumar P. A review on the municipal solid waste management status, challenges and potential for the future Indian cities. Environment, Development and Sustainability. 2023. vol. 25. pp. 13755–13803. DOI: 10.1007/s10668-022-02688-7.
2. Anttiroiko A.-V. Smart circular cities: Governing the relationality, spatiality, and digitality in the promotion of circular economy in an urban region. Sustainability. 2023. vol. 15. no. 17. pp. 1–41. DOI: 10.3390/su151712680.
3. Sanchez-Garcia E., Martinez-Falco J., Marco-Lajara B., Manresa-Marhuenda E. Revolutionizing the circular economy through new technologies: A new era of sustainable progress. Environmental Technology & Innovation. 2024. vol. 33. 103509 p. DOI: 10.1016/j.eti.2023.103509.
4. Kokate A., Jadhav T. A Novel Approach to EEG Artifact Removal Using ADASYN and Optimized Hierarchical 1D CNN. Informatics and Automation. 2025. vol. 24. no. 5. pp. 1408–1443. DOI: 10.15622/ia.24.5.6.
5. Szpilko D., de la Torre Gallegos A., Jimenez Naharro F., Rzepka A., Remiszewska A. Waste Management in the Smart City: Current Practices and Future Directions. Resources. 2023. vol. 12. no. 10. pp. 1–15.
6. Dahr J.M., Gaafar A.S. Performance Evaluation of the Convolutional Neural Networks for Object Identification Using RGB and Binary Images. Informatica. 2024. vol. 48. pp. 177–188. DOI: 10.31449/inf.v48i21.6568.
7. Rittl L.G.F., Zaman A., de Oliveira F.H. Digital Transformation in Waste Management: Disruptive Innovation and Digital Governance for Zero-Waste Cities in the Global South as Keys to Future Sustainable Development. Sustainability. 2025. vol. 17. no. 4. 1608 p. DOI: 10.3390/su17041608.
8. Fuqaha S., Nursetiawan N. Artificial intelligence and IoT for smart waste management: Challenges, opportunities, and future directions. Journal of Future Artificial Intelligence and Technology. 2025. vol. 2. no. 1. pp. 24–46.
9. Fotovvatikhah F., Ahmedy I., Noor R.M., Munir M.U. A systematic review of AI-based techniques for automated waste classification. Sensors. 2025. vol. 25. no. 10. 3181 p.
10. Fang B., Yu J., Chen Z., et al. Artificial intelligence for waste management in smart cities: a review. Environmental Chemistry Letters. 2023. vol. 21. pp. 1959–1989. DOI: 10.1007/s10311-023-01604-3.
11. Hoque M.J., Islam M.S., Khaliluzzaman M. AI-Powered Precision in Diagnosing Tomato Leaf Diseases. Complexity. 2025. vol. 2025. no. 1. pp. 1–21. DOI: 10.1155/cplx/7838841.
12. Mahmood M., Chowdhury P., Yeassin R., Hasan M., Ahmad T., Chowdhury N.-U.-R. Impacts of digitalization on smart grids, renewable energy, and demand response: An updated review of current applications. Energy Conversion and Management X. 2024. vol. 24. 100790 p.
13. Elahi M., Afolaranmi S.O., Martinez Lastra J.L., et al. A comprehensive literature review of the applications of AI techniques through the lifecycle of industrial equipment. Discover Artificial Intelligence. 2023. vol. 3. 43 p. DOI: 10.1007/s44163-023-00089-x.

14. Zhang Q., Zhang J., Yang S. Enhancing YOLOv8 Object Detection with Shape-IoU Loss and Local Convolution for Small Target Recognition. Informatica. 2025. vol. 49. no. 21. pp. 2–24.
15. Vukicevic A.M., Petrovic M., Jurisevic N., et al. Versatile waste sorting in small batch and flexible manufacturing industries using deep learning techniques. Scientific Reports. 2025. vol. 15. no. 1. pp. 1–11. DOI: 10.1038/s41598-025-87226-x.
16. Hoque M.J., Ahmed M.R., Uddin M.J., Faisal M.M.A. Automation of Traditional Exam Invigilation using CCTV and Bio-Metric. International Journal of Advanced Computer Science and Applications (IJACSA). 2020. vol. 11. no. 6. pp. 392–399.
17. Abdu H., Mohd Noor M.H. A Survey on Waste Detection and Classification using Deep Learning. IEEE Access. 2022. vol. 10. pp. 1–9. DOI: 10.1109/ACCESS.2022.3226682.
18. Olawade D.B., Fapohunda O., Wada O.Z., Usman S.O., Ige A.O., Ajisafe O., Oladapo B.I. Smart waste management: A paradigm shift enabled by artificial intelligence. Waste Management Bulletin. 2024. vol. 2. no. 2. pp. 244–263. DOI: 10.1016/j.wmb.2024.05.001.
19. Hoque M.J., et al. Incorporating Meteorological Data and Pesticide Information to Forecast Crop Yields Using Machine Learning. IEEE Access. 2024. vol. 12. pp. 47768–47786. DOI: 10.1109/ACCESS.2024.3383309.
20. Dao S.V.T., Le T.M., Tran H.M., Pham H.V., Vu M.T., Chu T. Integrating artificial intelligence for sustainable waste management: Insights from machine learning and deep learning. Watershed Ecology and the Environment. 2025. vol. 7. pp. 353–382.
21. Ali A.K.A.A., Aydın Y. Vision Transformer-Based Approach: A Novel Method for Object Recognition. Karadeniz Fen Bilimleri Dergisi. 2025. vol. 15. no. 1. pp. 560–576. DOI: 10.31466/kfbd.1620640.
22. Islam M.R., et al. Deep Learning and Computer Vision Techniques for Enhanced Quality Control in Manufacturing Processes. IEEE Access. 2024. vol. 12. pp. 121449–121479. DOI: 10.1109/ACCESS.2024.3453664.
23. Singh J., El-Sappagh S., Ali F., et al. Smart waste management: a systematic review and scientometric analysis of artificial intelligence applications. Environ Dev Sustain. 2025. DOI: 10.1007/s10668-025-05975-1.
24. Abdullah M., Abedin M.Z. Assessment of Plastic Waste Management in Bangladesh: A Comprehensive Perspective on Sorting, Production, Separation, and Recycling. Results in Surfaces and Interfaces. 2024. vol. 15. 100221 p. DOI: 10.1016/j.rsurfi.2024.100221.
25. Jangsamsi K. Conventional Machine Learning Approach for Waste Classification. Proceedings 6th Artificial Intelligence Cloud Computing Conference (AICCC). 2023. pp. 7–12.
26. Ramos E., Lopes A.G., Mendonça F. Application of Machine Learning in Plastic Waste Detection and Classification: A Systematic Review. Processes. 2024. vol. 12. no. 8. 1632 p. DOI: 10.3390/pr12081632.
27. Zhang Q., Yang Q., Zhang X., Bao Q., Su J., Liu X. Waste image classification based on transfer learning and convolutional neural network. Waste Management. 2021. vol. 135. pp. 150–157. DOI: 10.1016/j.wasman.2021.08.038.
28. Majchrowska S., et al. Deep learning-based waste detection in natural and urban environments. Waste Management. 2022. vol. 138. pp. 274–284. DOI: 10.1016/j.wasman.2021.12.001.
29. Azis F.A., Suhaimi H., Abas E. Waste Classification using Convolutional Neural Network. Proceedings 2nd International Conference on Information Technology, Computing, and Communication. 2020. pp. 9–13. DOI: 10.1145/3417473.3417474.
30. Tamin O., et al. On-Shore Plastic Waste Detection with YOLOv5 and RGB-Near-Infrared Fusion: A State-of-the-Art Solution for Accurate and Efficient Environmental

Monitoring. Big Data and Cognitive Computing. 2023. vol. 7. no. 2. pp. 1–27. DOI: 10.3390/bdcc7020103.
31. Oza P., et al. Solid waste classification using deep neural network: A transfer learning approach. Earth Science Informatics. 2025. vol. 18. 255 p. DOI: 10.1007/s12145-025-01743-x.
32. Liu C., Xie N., Yang X., Chen R., Chang X., Zhong R.Y., Peng S., Liu X. A Domestic Trash Detection Model Based on Improved YOLOX. Sensors. 2022. vol. 22. no. 18. 6974 p. DOI: 10.3390/s22186974.
33. Li J., Feng Y., Shao Y., Liu F. IDP-YOLOv9: Improvement of Object Detection Model in Severe Weather Scenarios from Drone Perspective. Applied Sciences. 2024. vol. 14. no. 12. 5277 p. DOI: 10.3390/app14125277.
34. Jabed M.R., Shamsuzzaman M. YOLObin: Non-Decomposable Garbage Identification and Classification Based on YOLOv7. Journal of Computer and Communications. 2022. vol. 10. no. 10. pp. 104–121. DOI: 10.4236/jcc.2022.1010008.
35. Raj T.S.P., et al. Classification of Waste with the Assistance of YOLO. Proceedings 11th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO). 2024. pp. 1–5.
36. Marelli A., Magri L., Arrigoni F., Boracchi G. Temporal-Consistent CAMs for Weakly Supervised Video Segmentation in Waste Sorting. In: Del Bue A., Canton C., Pont-Tuset J., Tommasi T. (eds). Computer Vision – ECCV 2024 Workshops. ECCV 2024. Lecture Notes in Computer Science. Cham: Springer. 2025. vol. 15626. pp. 371–387. DOI: 10.1007/978-3-031-92805-5_22.
37. Wang A., Liu L., Chen H., Lin Z., Han J., Ding G. YOLOE: Real-Time Seeing Anything. Proceedings of the International Conference on Computer Vision (ICCV). 2025. pp. 1–15. DOI: 10.48550/arXiv.2503.07465.
38. Zhang S., Zhu H., He Y., Guo M., Lou Z., Chang S. WISNet: Pseudo Label Generation on Unbalanced and Patch Annotated Waste Images. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2025. pp. 15076–15085.
39. Zhang S., Zhang L., Liu Z. Test-time adaptation for object detection via Dynamic Dual Teaching. Image and Vision Computing. 2025. vol. 163. 105740 p.
40. Proença P.F., Simoes P. TACO: Trash Annotations in Context for Litter Detection. arXiv preprint arXiv:2003.06975. 2020.
41. Bashkirova D., Zhu Z., Akl J., Alladkani F., Hu P., Ablavsky V., Calli B., Bargal S.A., Saenko K. ZeroWaste Dataset: Towards Automated Waste Recycling. Proceedings Conference on Neural Information Processing Systems (NeurIPS). 2021.
42. Nnamoko N., Barrowclough J., Procter J. Solid Waste Image Classification Using Deep Convolutional Neural Network. Infrastructures. 2022. vol. 7. no. 4. pp. 1–18. DOI: 10.3390/infrastructures7040047.
43. Shroff M., Desai A., Garg D. YOLOv8-based Waste Detection System for Recycling Plants: A Deep Learning Approach. Proceedings International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS). 2023. pp. 1–9. DOI: 10.1109/ICSSAS57918.2023.10331643.

Maiza Mohammed — Ph.D., Dr.Sci., Associate professor, LITIO Laboratory, Faculty of Exact and Applied Sciences, University of Oran 1 Ahmed Ben Bella (UO1). Research interests: bioinformatics, computational biology, optimization algorithms, digital image processing, machine learning, microarray data analysis, evolutionary computing, genetic programming, cancer genomics, feature selection in multidimensional biological data, cancer classification and early diagnosis systems. The number of publications — 7. maiza.mohammed@univ-oran1.dz; 1524, El M'Naouer, 31000, Oran, Algeria; office phone: +213(41)58-1947

Cherif Chahira — Ph.D., Dr.Sci., Associate professor, RIIR Laboratory, Faculty of Medicine, University of Oran 1 Ahmed Ben Bella (UO1). Research interests: artificial intelligence, business process management, software engineering, decision support systems, modeling of business rules, application of machine learning in healthcare and industry, optimization of business processes based on AI. The number of publications — 10. cherif.chahira@univ-oran1.dz; BP 1524, El M'Naouer, 31000, Oran, Algeria; office phone: +213(41)58-1947.

Chouraqui Samira — Ph.D., Dr.Sci., Professor, Faculty of Computer Science, University of Science and Technology of Oran - Mohamed Boudiaf (USTO-MB). Research interests: computer vision, artificial intelligence, unmanned aerial vehicles (UAVs), image recognition, application of machine learning in remote sensing, deep learning for image processing, AI-based solutions for environmental monitoring and precision farming. The number of publications — 65. samira.chouraqui@univ-usto.dz; BP 1505, El M'Naouer, 31000, Oran, Algeria; office phone: +213(41)56-0327.

Taleb-Ahmed Abdelmalik — Ph.D., Dr.Sci., Professor, Polytechnic University of Hauts-de-France (UPHF); Employee, Institute of Electronics, Microelectronics and Nanotechnology (IEMN). Research interests: computer vision, machine learning, pattern recognition, image segmentation and classification, data fusion, applications in biometrics, video surveillance, autonomous driving, medical imaging, deep learning for medical image analysis, multimodal data fusion. The number of publications — 534. abdelmalik.taleb-ahmed@uphf.fr; 9, Cedex, 59313, Valenciennes, France; office phone: 0(033)84-7392.

DOI 10.15622/ia.25.3.3

М. МАИЗА , С. ШЕРИФ , С. ШУРАКИ , А. ТАЛЕБ-АХМЕД

ТОЧНОЕ ОБНАРУЖЕНИЕ ГОРОДСКИХ ОТХОДОВ В РЕАЛЬНОМ ВРЕМЕНИ С ИСПОЛЬЗОВАНИЕМ УЛУЧШЕННОЙ АРХИТЕКТУРЫ YOLOV9-EFFICIENTVIT

Маиза М., Шериф С., Шураки С., Талеб-Ахмед А. **Точное обнаружение городских отходов в реальном времени с использованием улучшенной архитектуры YOLOv9-EfficientViT.**

Аннотация. Накопление небиоразлагаемых твердых отходов в густонаселенных городских районах представляет собой серьезную экологическую проблему. Хотя компьютерное зрение предлагает перспективное решение, современные подходы часто ограничены зависимостью от ограниченных синтетических данных и неспособностью охватить всю сложность реальных условий. Для решения этих ограничений в данной статье представлена система обнаружения объектов в реальном времени для городских неразлагаемых отходов, построенная на оптимизированной архитектуре YOLOv9. В основе работы лежит специально созданный гетерогенный набор данных, состоящий из 6340 аннотированных изображений, который охватывает восемь категорий отходов, включая пластиковые бутылки, банки и блистерные упаковки, в различных городских ландшафтах. Основное архитектурное улучшение заключается в замене исходного бэкбона на новый EfficientViT, который сочетает в себе преимущества многомасштабного извлечения признаков сверточных нейронных сетей с глобальным контекстным пониманием визуальных трансформеров. Такая гибридная архитектура особенно эффективна для обнаружения объектов отходов различных размеров в загроможденных городских средах. Дополнительные улучшения включают использование функции активации SiLU, масштабируемой функции потерь SIoU для точной регрессии ограничивающих рамок и фокальной функции потерь для противодействия дисбалансу классов. Модель, обученная с расширенной аугментацией данных, достигает mAP@0.5 95,1% и F1-меры 0,95 на отложенном тестовом наборе, превосходя все существующие детекторы отходов на основе YOLO. Все восемь классов достигают F1-меры выше 0,93, что свидетельствует о стабильной устойчивости системы даже в загроможденных средах. При работе со скоростью 38 кадров в секунду разработанная система подтверждает свою пригодность для практических применений в реальном времени. Используя передовые возможности YOLOv9 с современным бэкбоном EfficientViT и реалистичным набором данных, предлагаемая сеть устанавливает новый эталон точности и скорости в обнаружении отходов, демонстрируя большой потенциал для интеграции в автоматизированные системы сортировки и переработки.

Ключевые слова: обнаружение объектов, классификация отходов, компьютерное зрение, YOLOv9, EfficientViT, визуальный трансформер, реальное время, глубокое обучение.

Литература

1. Soni A., Das P.K., Kumar P. A review on the municipal solid waste management status, challenges and potential for the future Indian cities // Environment, Development and Sustainability. 2023. vol. 25. pp. 13755–13803. DOI: 10.1007/s10668-022-02688-7.
2. Anttiroiko A.-V. Smart circular cities: Governing the relationality, spatiality, and digitality in the promotion of circular economy in an urban region // Sustainability. 2023. vol. 15. no. 17. pp. 1–41. DOI: 10.3390/su151712680.
3. Sanchez-Garcia E., Martinez-Falco J., Marco-Lajara B., Manresa-Marhuenda E. Revolutionizing the circular economy through new technologies: A new era of sustainable

progress // Environmental Technology & Innovation. 2024. vol. 33. 103509 p. DOI: 10.1016/j.eti.2023.103509.

4. Kokate A., Jadhav T. A Novel Approach to EEG Artifact Removal Using ADASYN and Optimized Hierarchical 1D CNN // Informatics and Automation. 2025. vol. 24. no. 5. pp. 1408–1443. DOI: 10.15622/ia.24.5.6.
5. Szpilko D., de la Torre Gallegos A., Jimenez Naharro F., Rzepka A., Remiszewska A. Waste Management in the Smart City: Current Practices and Future Directions // Resources. 2023. vol. 12. no. 10. pp. 1–15.
6. Dahr J.M., Gaafar A.S. Performance Evaluation of the Convolutional Neural Networks for Object Identification Using RGB and Binary Images // Informatica. 2024. vol. 48. pp. 177–188. DOI: 10.31449/inf.v48i21.6568.
7. Rittl L.G.F., Zaman A., de Oliveira F.H. Digital Transformation in Waste Management: Disruptive Innovation and Digital Governance for Zero-Waste Cities in the Global South as Keys to Future Sustainable Development // Sustainability. 2025. vol. 17. no. 4. 1608 p. DOI: 10.3390/su17041608.
8. Fuqaha S., Nursetiawan N. Artificial intelligence and IoT for smart waste management: Challenges, opportunities, and future directions // Journal of Future Artificial Intelligence and Technology. 2025. vol. 2. no. 1. pp. 24–46.
9. Fotovvatikhah F., Ahmedy I., Noor R.M., Munir M.U. A systematic review of AI-based techniques for automated waste classification // Sensors. 2025. vol. 25. no. 10. 3181 p.
10. Fang B., Yu J., Chen Z., et al. Artificial intelligence for waste management in smart cities: a review // Environmental Chemistry Letters. 2023. vol. 21. pp. 1959–1989. DOI: 10.1007/s10311-023-01604-3.
11. Hoque M.J., Islam M.S., Khaliluzzaman M. AI-Powered Precision in Diagnosing Tomato Leaf Diseases // Complexity. 2025. vol. 2025. no. 1. pp. 1–21. DOI: 10.1155/cplx/7838841.
12. Mahmood M., Chowdhury P., Yeassin R., Hasan M., Ahmad T., Chowdhury N.-U.-R. Impacts of digitalization on smart grids, renewable energy, and demand response: An updated review of current applications // Energy Conversion and Management X. 2024. vol. 24. 100790 p.
13. Elahi M., Afolaranmi S.O., Martinez Lastra J.L., et al. A comprehensive literature review of the applications of AI techniques through the lifecycle of industrial equipment // Discover Artificial Intelligence. 2023. vol. 3. 43 p. DOI: 10.1007/s44163-023-00089-x.
14. Zhang Q., Zhang J., Yang S. Enhancing YOLOv8 Object Detection with Shape-IoU Loss and Local Convolution for Small Target Recognition // Informatica. 2025. vol. 49. no. 21. pp. 2–24.
15. Vukicevic A.M., Petrovic M., Jurisevic N., et al. Versatile waste sorting in small batch and flexible manufacturing industries using deep learning techniques // Scientific Reports. 2025. vol. 15. no. 1. pp. 1–11. DOI: 10.1038/s41598-025-87226-x.
16. Hoque M.J., Ahmed M.R., Uddin M.J., Faisal M.M.A. Automation of Traditional Exam Invigilation using CCTV and Bio-Metric // International Journal of Advanced Computer Science and Applications (IJACSA). 2020. vol. 11. no. 6. pp. 392–399.
17. Abdu H., Mohd Noor M.H. A Survey on Waste Detection and Classification using Deep Learning // IEEE Access. 2022. vol. 10. pp. 1–9. DOI: 10.1109/ACCESS.2022.3226682.
18. Olawade D.B., Fapohunda O., Wada O.Z., Usman S.O., Ige A.O., Ajisafe O., Oladapo B.I. Smart waste management: A paradigm shift enabled by artificial intelligence // Waste Management Bulletin. 2024. vol. 2. no. 2. pp. 244–263. DOI: 10.1016/j.wmb.2024.05.001.
19. Hoque M.J., et al. Incorporating Meteorological Data and Pesticide Information to Forecast Crop Yields Using Machine Learning // IEEE Access. 2024. vol. 12. pp. 47768–47786. DOI: 10.1109/ACCESS.2024.3383309.

20. Dao S.V.T., Le T.M., Tran H.M., Pham H.V., Vu M.T., Chu T. Integrating artificial intelligence for sustainable waste management: Insights from machine learning and deep learning // Watershed Ecology and the Environment. 2025. vol. 7. pp. 353–382.
21. Ali A.K.A.A., Aydın Y. Vision Transformer-Based Approach: A Novel Method for Object Recognition // Karadeniz Fen Bilimleri Dergisi. 2025. vol. 15. no. 1. pp. 560–576. DOI: 10.31466/kfbd.1620640.
22. Islam M.R., et al. Deep Learning and Computer Vision Techniques for Enhanced Quality Control in Manufacturing Processes // IEEE Access. 2024. vol. 12. pp. 121449–121479. DOI: 10.1109/ACCESS.2024.3453664.
23. Singh J., El-Sappagh S., Ali F., et al. Smart waste management: a systematic review and scientometric analysis of artificial intelligence applications // Environ Dev Sustain. 2025. DOI: 10.1007/s10668-025-05975-1.
24. Abdullah M., Abedin M.Z. Assessment of Plastic Waste Management in Bangladesh: A Comprehensive Perspective on Sorting, Production, Separation, and Recycling // Results in Surfaces and Interfaces. 2024. vol. 15. 100221 p. DOI: 10.1016/j.rsurfi.2024.100221.
25. Jangsamsi K. Conventional Machine Learning Approach for Waste Classification // Proceedings 6th Artificial Intelligence Cloud Computing Conference (AICCC). 2023. pp. 7–12.
26. Ramos E., Lopes A.G., Mendonça F. Application of Machine Learning in Plastic Waste Detection and Classification: A Systematic Review // Processes. 2024. vol. 12. no. 8. 1632 p. DOI: 10.3390/pr12081632.
27. Zhang Q., Yang Q., Zhang X., Bao Q., Su J., Liu X. Waste image classification based on transfer learning and convolutional neural network // Waste Management. 2021. vol. 135. pp. 150–157. DOI: 10.1016/j.wasman.2021.08.038.
28. Majchrowska S., et al. Deep learning-based waste detection in natural and urban environments // Waste Management. 2022. vol. 138. pp. 274–284. DOI: 10.1016/j.wasman.2021.12.001.
29. Azis F.A., Suhaimi H., Abas E. Waste Classification using Convolutional Neural Network // Proceedings 2nd International Conference on Information Technology, Computing, and Communication. 2020. pp. 9–13. DOI: 10.1145/3417473.3417474.
30. Tamin O., et al. On-Shore Plastic Waste Detection with YOLOv5 and RGB-Near-Infrared Fusion: A State-of-the-Art Solution for Accurate and Efficient Environmental Monitoring // Big Data and Cognitive Computing. 2023. vol. 7. no. 2. pp. 1–27. DOI: 10.3390/bdcc7020103.
31. Oza P., et al. Solid waste classification using deep neural network: A transfer learning approach // Earth Science Informatics. 2025. vol. 18. 255 p. DOI: 10.1007/s12145-025-01743-x.
32. Liu C., Xie N., Yang X., Chen R., Chang X., Zhong R.Y., Peng S., Liu X. A Domestic Trash Detection Model Based on Improved YOLOX // Sensors. 2022. vol. 22. no. 18. 6974 p. DOI: 10.3390/s22186974.
33. Li J., Feng Y., Shao Y., Liu F. IDP-YOLOv9: Improvement of Object Detection Model in Severe Weather Scenarios from Drone Perspective // Applied Sciences. 2024. vol. 14. no. 12. 5277 p. DOI: 10.3390/app14125277.
34. Jabed M.R., Shamsuzzaman M. YOLObin: Non-Decomposable Garbage Identification and Classification Based on YOLOv7 // Journal of Computer and Communications. 2022. vol. 10. no. 10. pp. 104–121. DOI: 10.4236/jcc.2022.1010008.
35. Raj T.S.P., et al. Classification of Waste with the Assistance of YOLO // Proceedings 11th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO). 2024. pp. 1–5.

36. Marelli A., Magri L., Arrigoni F., Boracchi G. Temporal-Consistent CAMs for Weakly Supervised Video Segmentation in Waste Sorting / In: Del Bue A., Canton C., Pont-Tuset J., Tommasi T. (eds) / Computer Vision – ECCV 2024 Workshops. ECCV 2024. Lecture Notes in Computer Science. Cham: Springer. 2025. vol. 15626. pp. 371–387. DOI: 10.1007/978-3-031-92805-5_22.
37. Wang A., Liu L., Chen H., Lin Z., Han J., Ding G. YOLOE: Real-Time Seeing Anything // Proceedings of the International Conference on Computer Vision (ICCV). 2025. pp. 1–15. DOI: 10.48550/arXiv.2503.07465.
38. Zhang S., Zhu H., He Y., Guo M., Lou Z., Chang S. WISNet: Pseudo Label Generation on Unbalanced and Patch Annotated Waste Images // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2025. pp. 15076–15085.
39. Zhang S., Zhang L., Liu Z. Test-time adaptation for object detection via Dynamic Dual Teaching // Image and Vision Computing. 2025. vol. 163. 105740 p.
40. Proença P.F., Simoes P. TACO: Trash Annotations in Context for Litter Detection // arXiv preprint arXiv:2003.06975. 2020.
41. Bashkirova D., Zhu Z., Akl J., Alladkani F., Hu P., Ablavsky V., Calli B., Bargal S.A., Saenko K. ZeroWaste Dataset: Towards Automated Waste Recycling // Proceedings Conference on Neural Information Processing Systems (NeurIPS). 2021.
42. Nnamoko N., Barrowclough J., Procter J. Solid Waste Image Classification Using Deep Convolutional Neural Network // Infrastructures. 2022. vol. 7. no. 4. pp. 1–18. DOI: 10.3390/infrastructures7040047.
43. Shroff M., Desai A., Garg D. YOLOv8-based Waste Detection System for Recycling Plants: A Deep Learning Approach // Proceedings International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS). 2023. pp. 1–9. DOI: 10.1109/ICSSAS57918.2023.10331643.

Майза Мохаммед — д-р филос. наук, доцент, лаборатория LITIO, факультет точных и прикладных наук, Университет Орана 1 Ахмед Бен Белла (УО1). Область научных интересов: биоинформатика, вычислительная биология, алгоритмы оптимизации, цифровая обработка изображений, машинное обучение, анализ данных микрочипов, эволюционные вычисления, генетическое программирование, геномика рака, отбор признаков в многомерных биологических данных, классификация рака и системы ранней диагностики. Число научных публикаций — 7. maiza.mohammed@univ-oran1.dz; Эль-М'Науэр, ВР 1524, 31000, Оран, Алжир; р.т.: +213(41)58-1947.

Шериф Чахира — д-р филос. наук, доцент, лаборатория RIIR, медицинский факультет, Университет Орана 1 Ахмед Бен Белла (УО1). Область научных интересов: искусственный интеллект, управление бизнес-процессами, программная инженерия, системы поддержки принятия решений, моделирование бизнес-правил, применение машинного обучения в здравоохранении и промышленности, оптимизация бизнес-процессов на основе ИИ. Число научных публикаций — 10. cherif.chahira@univ-oran1.dz; Эль-М'Науэр, ВР 1524, 31000, Оран, Алжир; р.т.: +213(41)58-1947.

Шураки Самира — д-р филос. наук, профессор, факультет компьютерных наук, Университет науки и технологии Оран имени Мохамеда Будиафа (УСТО-МБ). Область научных интересов: компьютерное зрение, искусственный интеллект, беспилотные летательные аппараты (БПЛА), распознавание образов, применение машинного обучения в дистанционном зондировании, глубокое обучение для обработки изображений, решения на основе ИИ для мониторинга окружающей среды и точного земледелия. Число научных публикаций — 65. samira.chouraqui@univ-usto.dz; Эль-М'Науэр, ВР 1505, 31000, Оран, Алжир; р.т.: +213(41)56-0327.

Талеб-Ахмед Абдельмалик — д-р филос. наук, профессор, Политехнический университет О-де-Франс (UPHF); сотрудник, Институт электроники, микроэлектроники и нанотехнологий (ИЕМН). Область научных интересов: компьютерное зрение, машинное обучение, распознавание образов, сегментация и классификация изображений, слияние данных, применение в биометрии, видеонаблюдении, автономном вождении, медицинской визуализации, глубокое обучение для анализа медицинских изображений, мультимодальное слияние данных. Число научных публикаций — 534. abdelmalik.taleb-ahmed@uphf.fr; Седекс, 9, 59313, Валансьен, Франция; р.т.: 0(033)84-7392.