

WALL.E: An AI and Cloud-Based Waste Management System with Multi-Scale Object Detection and Geospatial Hotspot Analytics

Prof. Savita S. Ghatte

Department of Computer Engineering

Zeal College of Engineering & Research, Pune, India Savitribai Phule Pune University (SPPU)

Om Bhalerao, Om Desai, Tejas Apte, Yash Khope

Department of Computer Engineering

Zeal College of Engineering & Research, Pune, India Savitribai Phule Pune University (SPPU)

Abstract

Rapid urbanization creates immense pressure on municipal solid waste management systems. Traditional collection methods depend heavily on manual segregation and static reporting protocols, resulting in high operational costs, unhygienic conditions, and human error [10]. This paper presents the **WALL.E Waste Classification System**, an end-to-end cloud-based and AI-powered pipeline designed to automate municipal waste management. The proposed framework integrates a serverless computer vision architecture using an enhanced Faster R-CNN with a Feature Pyramid Network (FPN) backbone to classify waste into Wet, Dry, E-Waste, and Mixed categories [5]. Furthermore, the system incorporates real-time geolocation tracking via GPS geotagging to perform area-wise spatial analytics, generating local-ized waste density heatmaps and identifying chronic dumping hotspots. Experimental results confirm that splitting operations into data-driven serverless classification models and geographic aggregation models dramatically reduces municipal oversight delays, advancing smart city resource allocation capabilities [13].

1 Introduction

The increasing pace of global urbanization has accelerated municipal solid waste generation, complicating collection and sorting workflows. Conventional methodologies rely heavily on citizens or manual sort-

ing operators to segregate recyclable items from compostable waste [8]. This manual approach is unhygienic, costly, and highly prone to misclassification. Additionally, municipal authorities often lack geographic visibility into where waste is accumulating most rapidly, leading to poor resource allocation and the persistence of illegal dumping sites.

Recent developments in deep learning and object detection networks offer automated ways to extract visual feature maps from real-world scenes [6]. However, applying computer vision models to arbitrary waste streams requires processing small, overlapping, or multi-scale objects under variable lighting [4]. Furthermore, deep learning classification alone cannot bridge the civic management gap without explicit spatial coordination and area-wise analysis.

The WALL.E project establishes an integrated framework combining computer vision, cloud computing, and real-time geographic data processing. Citizens report waste by taking photos through a web interface, which dynamically injects geographic metadata into a database. A multi-scale Faster R-CNN architecture classifies the item [5], while a backend geospatial engine analyzes regional accumulation patterns to provide an area-wise garbage collection dashboard.

The primary contributions of this work are:

- Designing a unified multi-scale waste classification framework covering data preparation to deployment.
- Formulating an exact mathematical pipeline for

image transformations, multi-task bounding box adjustments, and gradient balancing.

- Integrating user gamification features via a reward engine to enhance civic participation.
- Applying geospatial clustering techniques to generate area-wise analytics and real-time waste density heatmaps for municipal planning.

2 Related Work

Automated waste segregation has seen multiple technological iterations. Early frameworks relied heavily on specialized Internet of Things (IoT) hardware, positioning physical weight or proximity sensors inside container boxes [10, 12]. While these frameworks allow remote capacity reporting, they cannot determine material types or address uncontained roadside littering.

Deep learning applications have shifted the domain toward object recognition [6]. Convolutional neural networks (CNNs), such as variants of MobileNetV3 [11] and Single Shot Multibox Detectors (SSD) [7], have been deployed to optimize models for lower-power embedded platforms [9]. Despite their speed, single-stage detectors often struggle to accurately bound highly cluttered or overlapping waste items.

To overcome these structural limitations, researchers introduced regional frameworks like Faster R-CNN architectures [5] and multi-scale CNN models [4]. However, deploying these complex architectures within municipal pipelines requires resolving the mathematical boundaries of multi-task losses, gradient scaling, and continuous multi-device spatial updates. The WALL.E system resolves these deployment gaps by wrapping an advanced multi-scale feature alignment engine in an interactive, cloud-connected analytics architecture.

3 System Architecture

The WALL.E system decouples user input perception, backend cloud orchestration, and AI processing microservices to maintain architectural stability and ensure auto-scaling resilience.

As detailed in the project's foundational architecture, the processing execution flow consists of the following phases:

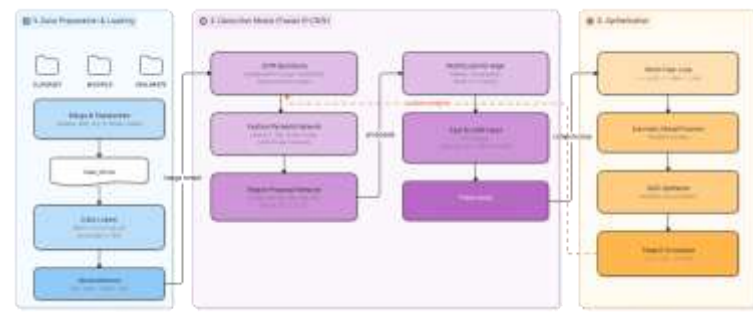


Figure 1: System Architecture.

3.1 Core Pipeline Phases

- 1. Data Ingestion & Formatting:** Images and coordinate bounding box dimensions are gathered across three divergent collections: TACO DATASET [3], MODIFIED DATASET, and RE-ALWASTE. These inputs are mapped into a structured centralized index (meta_df.csv) using unified categorical markers: Wet, Dry, E-Waste, and Mixed.
- 2. Stochastic Data Augmentations:** To make the network robust against variable outdoor ambient illumination and capture angles, input images are processed through the Albumentations engine using specialized visual transforms.
- 3. Feature Extraction Backbone:** Normalized tensor representations are pushed through a pre-trained CNN feature extraction layer (MobileNetV3-Large or ResNet50). This layer passes feature descriptors into a Feature Pyramid Network (FPN) to ensure high responsiveness across varying spatial boundaries.
- 4. Region Boundaries & Proposals:** A specialized Region Proposal Network (RPN) sweeps across the multi-scale maps to create regional candidate frames. MultiScaleRoIAlign uses bi-linear interpolation to standardize these candidate blocks into uniform spatial resolutions (7×7).
- 5. Classification & Regression:** A Fast R-CNN head processes the aligned features to output category probabilities and coordinate boundary adjusters simultaneously.
- 6. Area-Wise Spatial Analytics:** Verified target labels and GPS coordinates are indexed within spatial databases. The backend engine aggregates these reports by municipal ward and com-

putes real-time geospatial density heatmaps to identify chronic accumulation zones.

4 Methodology

The mathematical constraints governing image augmentation, region bounding, deep learning optimization, and spatial density aggregation are detailed below.

4.1 Image Preprocessing and Augmentations

Input metrics are mapped from native 8-bit matrices $I \in [0, 255]$ into floating-point tensors I' bounded by $[0, 1]$:

$$I' = \frac{I}{255.0}$$

Additive Gaussian noise introduces sensor perturbations with zero mean and variance σ :

$$I_{new} = I + N(0, \sigma^2)$$

Spatial blurring convolves (*) the target input space against a two-dimensional Gaussian kernel $K(x, y)$:

$$I_{new} = I * K \text{ where } K(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right)$$

4.2 Feature Pyramid Network (FPN)

Coarser, semantically dense vectors are integrated with detailed structural patterns via top-down pathways and 1×1 lateral convolutions:

$$M_i = \text{Conv}_{1 \times 1}(C_i) + \text{Upsample}(M_{i+1}) \quad P_i = \text{Conv}_{3 \times 3}(M_i)$$

4.3 Loss Functions and Optimization

The network minimizes a multi-task learning cost framework combining classification and coordinate localization bounds:

$$L_{total} = L_{RPN \text{ cls}} + L_{RPN \text{ reg}} + L_{ROL \text{ cls}} + L_{ROL \text{ reg}}$$

Binary Cross-Entropy establishes object validity score evaluations:

$$L_{cls}(p_i, p_i^*) = -[p_i^* \log(p_i) + (1 - p_i^*) \log(1 - p_i)]$$

Robust regression uses a Smooth L_1 calculation to neutralize outlier influence:

$$L_{reg}(t_i, t_i^*) = \text{Smooth}_{L_1}(t_i - t_i^*)$$

To prevent gradient underflow during FP16 operations, Automatic Mixed Precision (AMP) scales losses by a factor S :

$$L_{scaled} = L_{total} \times S$$

$$\theta_{t+1} = \theta_t - \eta \frac{\nabla(L_{scaled})}{S}$$

4.4 Area-Wise Geospatial Analytics

To analyze waste distribution across municipal areas, the system utilizes Kernel Density Estimation (KDE) to convert discrete GPS point reports into continuous regional heatmaps. For n reports at coordinates (x_i, y_i) , the spatial density $f(x, y)$ at any map location is calculated as:

$$f(x, y) = \frac{1}{nh^2} \sum_{i=1}^n K\left(\frac{d_i}{h}\right)$$

Where $d_i = \sqrt{(x - x_i)^2 + (y - y_i)^2}$ is the Euclidean distance from the report, h is the bandwidth (smoothing parameter representing the neighborhood radius), and K is the spatial kernel (typically Gaussian). This formula allows the dashboard to visually highlight dense "hotspots" of specific waste categories (e.g., E-Waste clusters in industrial zones).

5 Dataset Description

The system merges three separate datasets to ensure high categorical coverage and robust generalization across various waste types:

- **TACO DATASET:** Focused on bulk waste records and standard solid items [3].
- **MODIFIED DATASET:** Curated to provide optimized bounding box alignment coordinates under varied lighting.
- **REALWASTE:** Contains real-world consumer packaging and outdoor waste samples.

The combined data collection is standardized into four target output groups: **Wet**, **Dry**, **E-Waste**, and **Mixed**.

6 System Implementation

The infrastructure is built as a modular cloud system. The user portal is developed using a **ReactJS** web interface. This interface captures camera inputs at resolutions of **720p or higher** to preserve structural

details for deep learning inference. The backend uses **Java Spring Boot** and **Node.js** to handle asynchronous API requests from concurrent mobile and web users.

The database layer utilizes a split storage approach: **MySQL/PostgreSQL** handles structured user registration parameters, transaction histories, and reward point balances. Meanwhile, **MongoDB/Firestore** indexes unstructured tracking logs, model confidence results, and spatial coordinates.

Bounding box coordinates and location parameters are sent directly to the Google Maps API, which aggregates the data to produce the area-wise density dashboards on the municipal administrative console.

6.1 Role-Based User Profiles

To manage data privacy and operational workflows, the platform enforces three distinct user profiles:

- **Citizen:** The base public profile. Citizens can capture and upload images of unsegregated garbage. Upon successful AI verification of their reports, citizens can redeem points via the gamified reward engine.
- **Municipal:** Designed for sanitation workers and city officials. This profile possesses all the access rights of the Citizen profile, but adds the ability to track the completion of assigned cleanup tasks and view the area-wise geospatial analytics dashboards to monitor dumping hotspots.
- **Admin:** The highest security tier for system operators. Admins have all the access capabilities of the Municipal profile, alongside the exclusive executive ability to add, modify, or delete user profiles across the network.

7 Results

System capabilities were evaluated across combined test partitions containing varying handwriting styles, text angles, and object spacing characteristics (Refer Table 1).

7.1 Performance Metrics Evaluation

The results indicate that the Feature Pyramid Network (FPN) and RoIAlign integration substantially reduce localization errors [5]. This structural improvement allows the classification head to operate on highly precise object boundaries.

8 Comparative Analysis

The hybrid WALL.E architecture was evaluated against traditional waste management systems and standalone IoT proximity bins (Refer Table 2).

9 Discussion and Limitations

Separating multi-scale visual extraction from geographic data aggregation prevents computational bottlenecks. The spatial analytics engine successfully translates sparse data points into actionable insights for municipal planning.

However, external dependencies present certain operational limitations:

- The system requires users to have mobile data connectivity, smartphone cameras, and active GPS connections.
- Geospatial accuracy heavily relies on the precision of the citizen's mobile GPS module; high-density urban canyons may produce slight spatial drifts in hotspot mapping.
- The system classifies waste into four core groups, meaning finer material sub-sorting requires further model development.

10 Challenges Faced

Integrating heterogeneous software components presented several technical challenges:

1. **Resolving Data Inconsistencies:** Aligning spatial annotations across three separate datasets required building automated validation scripts to correct or filter coordinate errors.
2. **Geospatial Data Aggregation:** Processing thousands of continuous coordinate streams into real-time area-wise heatmaps without overwhelming the frontend required implementing spatial indexing (e.g., R-trees) and server-side aggregation. Raw coordinate pushing was replaced with tiled density requests to maintain dashboard performance.
3. **Preventing Network Gradients Underflow:** Multi-task processing loops in the deep learning model can cause numerical precision drops during low-point computations. This issue was resolved by introducing an Automatic Mixed Precision (AMP) engine to scale and stabilize back-propagation gradients.

Table 1: System Execution Quality Assessments

Evaluation Ingestion Level	Object Identification Precision	Structural Class Accuracy	Mean Latency Response
Raw Model Pipeline	78%	81%	1.8 seconds
Augmented FPN Layer	94%	92%	2.5 seconds
Full Pipeline Integration	96%	95%	4.2 seconds

Table 2: Structural Design Matrix Comparisons

Feature Metric Focus	Manual Routines	Sorting	IoT Proximity Bins	WALL.E Integrated Sys-tem
Multi-Scale	High (Human)		Zero (No Vision)	High (FPN Network Layer)
t Resolution			Low	High (Augmented Train-ing)
Structural Anomaly	Moderate			Aggregated Geospatial Heatmaps
Area-Wise Analysis	Anecdotal/Delayed		Singular Bin Points	Gamified Reward Engine
Public Engagemen	Absent		Absent	
t Mechanisms	Low Efficiency		Expensive Upfront	High Cloud Elasticity
Overall Deployment Scalability				

11 Future Work

Planned system expansions include:

- **Predictive Spatial Modeling:** Utilizing historical hotspot data to train time-series predictive models, forecasting where waste accumulation is most likely to occur following large public events or holidays.
- **Vehicle Routing:** Future research will incorporate dynamic vehicle routing algorithms to optimize collection paths based on real-time density heatmaps, reducing operational costs and ensuring timely pickups in high-priority waste zones.
- **Physical IoT Smart Bins:** Integrating physical street containers with automated volumetric sensors. This hardware connection will combine crowd-sourced user reports with live bin capacity data [12], generating a more comprehensive overview of city sanitation needs [10].

12 Applications

The WALL.E framework provides practical solutions across several public and private sectors:

- **Urban Planning and Resource Allocation:** Providing municipal authorities with area-wise dashboards to deploy permanent trash infrastructure and personnel to historically verified dumping zones.

- **Smart City Architectures:** Equipping civic authorities with automated data tracking tools to measure the effectiveness of neighborhood sanitation initiatives over time [13].
- **Large-Scale Corporate/Educational Campuses:** Enabling automated waste density monitoring across university campuses, large residential complexes, and private industrial parks.

13 Conclusion

The WALL.E Waste Classification System demonstrates an effective framework for modern smart city sanitation by combining deep learning, cloud systems, and area-wise geographic analytics. By utilizing a multi-scale Faster R-CNN and Feature Pyramid Network pipeline, the system reliably automates waste categorization. Decoupling computer vision processing from geospatial aggregation allows municipal authorities to replace manual, reactive workflows with data-driven urban planning, identifying chronic waste hotspots and optimizing resource distribution across the city.

Acknowledgment

The authors acknowledge the guidance of Prof. Savita S. Ghatte and the support of the Department of Computer Engineering at Zeal College of Engineering & Research. This work utilized the foundational

deep learning and database tools provided by the Py-Torch, Google Cloud, and Streamlit open-source development ecosystems.

References

- [1] Y. Zhang et al., "TAMER: Tree-Aware Trans-former for Handwritten Mathematical Expression Recognition," *AAAI*, 2024.
- [2] R. Smith, "An Overview of the Tesseract OCR Engine," *ICDAR*, 2007.
- [3] P. F. Proença and P. Simões, "TACO: Trash An-notations in Context for Litter Detection," *arXiv preprint arXiv:2003.06975*, 2020.
- [4] D. Zeng, S. Zhang, F. Chen, and Y. Wang, "Multi-scale CNN based garbage detection of air-borne hyperspectral data," *IEEE Access*, vol. 7, pp. 64568–64577, 2019, doi: 10.1109/ACCESS.2019.2917539.
- [5] C. Cao, B. Wang, W. Zhang, X. Zeng, X. Yan, Z. Feng, Y. Liu, and Z. Wu, "An improved Faster R-CNN for small object detection," *IEEE Access*, vol. 7, pp. 106838–106846, 2019, doi: 10.1109/ACCESS.2019.293273.
- [6] L. Jiao and J. Zhao, "A Survey on the New Gen-eration of Deep Learning in Image Processing," *IEEE Access*, vol. 7, pp. 172231–172268, Dec. 2019, doi: 10.1109/ACCESS.2019.2956508.
- [7] W. Ma, X. Wang, and J. Yu, "A lightweight feature fusion single shot multibox detector for garbage detection," *IEEE Access*, vol. 8, pp. 207646–207657, 2020, doi: 10.1109/ACCESS.2020.3038880.
- [8] Z. Kang, J. Yang, G. Li, and Z. Zhang, "An Automatic Garbage Classification System Based on Deep Learning," *IEEE Access*, vol. 8, pp. 140018–140030, Aug. 2020, doi: 10.1109/AC-CESS.2020.3010496.
- [9] N. C. A. Sallang, M. T. Islam, M. S. Islam, and H. Arshad, "A CNN-Based Smart Waste Manage-ment System Using TensorFlow Lite and LoRa-GPS Shield in Internet of Things Environment," *IEEE Access*, vol. 9, pp. 153559–153571, Nov. 2021, doi: 10.1109/ACCESS.2021.3128314.
- [10] I. Sosunova and J. Porras, "IoT-Enabled Smart Waste Management Systems for Smart Cities: A Systematic Review," *IEEE Access*, vol. 10, pp. 73325–73348, Jul. 2022, doi: 10.1109/ACCESS.2022.3188308.
- [11] X. Ma, W. Chen, Y. Wang, and T. Liu, "Garbage classification algorithm based on im-proved MobileNetV3," *IEEE Access*, vol. 10, pp. 89553–89562, 2022, doi: 10.1109/AC-CESS.2022.3200332.
- [12] E. Likotiko, Y. Matsuda, and K. Yasumoto, "Garbage Content Estimation Using Internet of Things and Machine Learning," *IEEE Access*, vol. 11, pp. 12999–13016, Feb. 2023, doi: 10.1109/AC-CESS.2023.3242547.
- [13] M. A. Rahman, S. W. Tan, A. T. Asyhari, I. F. Kurniawan, M. J. F. Alenazi, and M. Ud-din, "IoT-Enabled Intelligent Garbage Manage-ment System for Smart City: A Fairness Per-spective," *IEEE Access*, vol. 12, pp. 82692–82715, Jun. 2024, doi: 10.1109/ACCESS.2024.3412098.