ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XII-4/W2-2026
10th International Conference on Smart Data and Smart Cities (SDSC), 29 September–1 October 2026, Sofia, Bulgaria

Zero-Shot Detection for Automatic Mapping of Illegal Roadside Waste Dumps from Volunteered Street-View Imagery

Al Maimun As Samee, Lukas Arzoumanidis, Huynh Duc An Son Nguyen, Youness Dehbi

Computational Methods Lab, HafenCity University, Hamburg, Germany -
{al.samee, lukas.arzoumanidis, son.nguyen, youness.dehbi}@hcu-hamburg.de

Keywords: zero-shot detection, roadside waste, street-view imagery, waste management, decision-support system.

Abstract

Illegal roadside waste dumping represents a persistent urban environmental problem, particularly in rapidly growing cities. Conventional monitoring approaches based on manual field surveys and mapping are labor-intensive, costly, and difficult to scale. This study presents an automated framework for detecting and mapping illegal roadside waste using volunteered street-view imagery and zero-shot semantic segmentation with the Segment Anything Model 3 (SAM 3). A dataset of approximately 14,000 Mapillary images from Dhaka, Bangladesh, collected between 2023 and 2025, was used for the analysis. Waste-related objects were identified using SAM 3 with text prompts, eliminating the need for task-specific training data. Detected instances were geolocated and visualized through an interactive web-based application and can be exported in GeoJSON format for further analysis outside of our web-based application. The results demonstrate the potential of combining visual foundation models with street-level imagery for scalable and cost-effective urban waste monitoring in resource-constrained settings. The proposed workflow is transferable to other cities with sufficient street-level imagery coverage and can support evidence-based municipal planning, cleanup operations, and enforcement. The developed code is publicly available at: `https://github.com/hcu-cml/roadside-waste-detection-mapping`.

1. Introduction

Illegal roadside waste dumping has become a pressing environmental challenge in rapidly urbanizing cities of developing countries, as municipal solid waste management infrastructure struggles to keep pace with population growth and waste generation rates (Wikurendra et al., 2024; Ngalo and Thondhlana, 2023). Poor waste management leads to environmental degradation and public health hazards, including contamination of soil and groundwater, respiratory complications (Wikurendra et al., 2024; Ichipi and Senekane, 2023) and may further increase the risk of urban flooding when drainage systems are obstructed by the accumulation of non-biodegradable roadside waste (Zhang et al., 2024).

Together with stronger policy implementation and stricter legislation, systematic roadside waste monitoring is an important component of improving waste management in rapidly urbanizing cities (Du et al., 2023). One reason is the limited documentation and spatial tracking of dumping sites, which hinders systematic analysis of temporal trends, spatial patterns, and causal factors that could inform preventive strategies and targeted enforcement efforts (Fazzo et al., 2020). Consequently, such monitoring enables the identification of areas with underdeveloped waste management infrastructure, helping stakeholders determine suitable locations for organized waste disposal sites (Zhang et al., 2025). Furthermore, since many roadside waste sites persist for extended periods before being discovered and remediated, sanitation workers can benefit from knowing the precise spatial locations of waste. This reduces the need for continuous visual searching and facilitates more efficient collection.

Traditional approaches to monitoring illegal dumping sites rely primarily on manual surveys, citizen complaints, and periodic field inspections by municipal staff (Cicala et al., 2024). These methods are inherently time-consuming, expensive, and difficult to scale across extensive urban areas, resulting in incomplete spatial coverage and delayed detection of emerging dumping incidents (Kim and Cho, 2022). In contrast, volunteered street-view imagery (SVI) has an extensive geographic coverage across diverse urban environments (Ma et al., 2026). Consequently, SVI represents a widely accessible and scalable data source for urban analysis and monitoring tasks. However, previous approaches to analyzing visual pollution and detecting roadside waste have largely relied on supervised deep learning methods, such as the YOLO architecture (Hossain et al., 2021). These approaches have limited applicability and transferability because of the substantial variation in urban morphology and waste appearance across different cityscapes, as well as the corresponding need for large amounts of annotated training data.

As a result, we present an approach for a location-based decision-support system for automated city-wide monitoring and mapping of roadside waste using SVI and a zero-shot visual foundation model, demonstrated through a case study in Dhaka, Bangladesh. The proposed approach is flexible and potentially transferable to other cities due to the zero-shot capabilities of visual foundation models. To improve usability, we developed a web-based application that visualizes areas with aggregated roadside waste using a heatmap representation on a two-dimensional (2D) map, as illustrated in Figure 1. Additionally, our approach provides the GPS coordinates of detected roadside waste, together with the corresponding image acquisition timestamps, in GeoJSON format. This ensures usability beyond the web-based visualization and facilitates further spatial and temporal analyses.

The remainder of this paper is structured as follows: Section 2 reviews recent advances in the prediction and mapping of roadside waste while Section 3 describes our data sources. Section 4 presents our proposed methodology in detail. Section 5 evaluates the quality and demonstrates the effectiveness of our approach. Finally, Section 6 concludes the paper and outlines directions for future research.

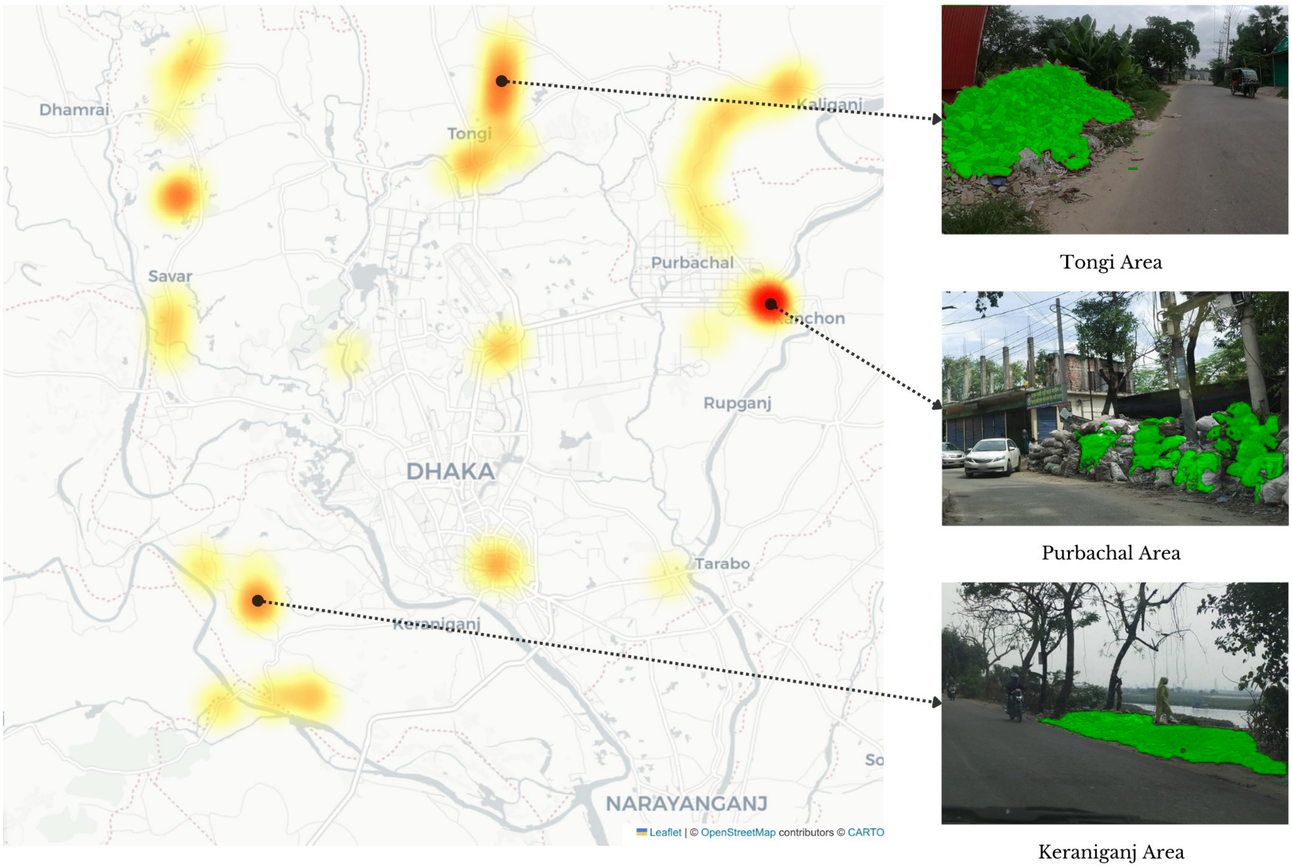


Figure 1. Heatmap visualizing the spatial density distribution of large quantities of roadside waste within the web application. The web application can be found under: `https://www.cml.hcu-hamburg.de/demos/roadside-waste-detection-mapping/`. Green segmentation mask within the SVIs shows detected roadside waste for exemplary Mapillary images. Lighter colors indicate lower roadside waste density, whereas darker colors indicate higher roadside waste density.

2. Related Work

Automated detection of illegally dumped waste has been investigated using a range of remote sensing and computer vision approaches, progressing from traditional machine learning image classification methods to deep learning-based techniques and, more recently, to foundation models with zero-shot generalization capabilities.

High-resolution satellite and aerial imagery have been widely used to identify dumpsites at both urban and global scales. In this context, deep convolutional networks based on Faster R-CNN and Feature Pyramid Networks (FPN) (Lin et al., 2017) have been applied to dumpsite detection across 28 cities, reducing investigation time by more than 96% compared with manual photo interpretation and enabling the analysis of social drivers underlying dumpsite distribution (Sun et al., 2023). In addition, semantic segmentation architectures such as DeepLabV3+, PSPNet, and HRNet have been employed on remote sensing imagery for waste-related detection tasks (Yong et al., 2023). Earlier work formulated illegal landfill identification as a multi-scale aerial scene-classification problem using a deep learning-based ResNet50 architecture (Torres and Fraternali, 2021). More recently, large vision foundation models such as Meta's Segment Anything Model (SAM) (Kirillov et al., 2023) have demonstrated strong zero-shot segmentation capabilities in remote sensing applications, highlighting their potential for detection and segmentation tasks in this domain (Osco et al., 2023).

Although satellite and aerial imagery support large-scale dumpsite mapping, they remain fundamentally constrained by occlusions, tree canopy cover, and limited spatial resolution, which hinder the reliable detection of small or partially concealed roadside waste accumulations. These smaller sites often constitute the majority of illegal dumping in dense urban environments. Street-level imagery, captured from ground-level viewpoints, directly reflects the visual perspective from which such waste is typically observable and therefore provides complementary spatial detail that aerial imagery cannot capture (Biljecki and Ito, 2021). To date, however, most studies have focused on waste site detection from aerial or satellite imagery, while only a limited number have explored street-view imagery for this purpose.

Volunteered street-level imagery platforms, particularly Mapillary, have emerged as promising open alternatives to commercial services such as Google Street View. Unlike proprietary platforms, Mapillary operates under a free license and accepts user-contributed imagery collected with consumer devices. Mapillary data have been used in a variety of urban monitoring applications, including the evaluation of street-level greenness (Zheng and Amemiya, 2023) and the prediction of building storey number (Arzoumanidis et al., 2026), demonstrating the platform's utility for automated urban analytics.

https://doi.org/10.5194/isprs-annals-XII-4-W2-2026-9-2026 | © Author(s) 2026. CC BY 4.0 License.

Nevertheless, the use of Mapillary imagery is associated with several important challenges. Spatial coverage is often inconsistent (Fan et al., 2025), while the imagery is further affected by occlusions, distortions, varying viewing angles, and substantial visual noise (Hou and Biljecki, 2022). These limitations make the detection of illegal dumping in volunteered street-level imagery challenging.

Large language models (LLMs), for example, have been employed to extract latent semantic representations from images of waste sites (Zhang et al., 2025). Related work has also integrated visual indicators of potential roadside dumping with socioeconomic variables derived from census data to support more robust detection and interpretation of waste-site occurrence (Yang et al., 2025). Furthermore, transformer-based models, including the Swin Transformer architecture (Liu et al., 2021), have been applied to automated waste detection in street-view imagery, facilitating the analysis of associations between urban environmental conditions and socioeconomic characteristics (Zhang and Ma, 2026). Collectively, these studies reflect a broader methodological transition toward multimodal, context-sensitive frameworks for understanding illegal dumping in urban environments.

In addition to illegal roadside waste detection, the automated identification of garbage bins and waste containers has also been investigated using deep learning-based object detection methods, with most studies relying on variants of the YOLO architecture (Li and Dai, 2026; Dong et al., 2025; Moral et al., 2022). In response to the challenges of using street-view imagery for waste detection, Paulo et al. (2026) proposed a multi-task dataset for urban waste management. This dataset is designed to support three core evaluation tasks, namely, waste container detection, waste container tracking, and waste overflow segmentation.

3. Study Area & Data Acquisition

Dhaka, Bangladesh, is among the most densely populated megacities worldwide and produces an estimated 3,500 tons of waste per day, with more than 12% of the generated waste being dumped along roadsides and in open spaces (Bahauddin and Uddin, 2012). Owing to the relatively good spatial coverage of Mapillary imagery, the city was selected as the area of interest for demonstrating the effectiveness of the proposed approach. The study area defined by the bounding box coordinates 90.25°E to 90.55°E longitude and 23.65°N to 23.95°N latitude, covers approximately $900\,\text{km}^2$. Street-view imagery was obtained from Mapillary and provides geotagged street-level images through its Application Programming Interface (API).

The Mapillary API was queried to retrieve images within the defined study area. Temporal filtering was applied to retain only images captured after January 1, 2023, ensuring a contemporary representation of urban conditions. Due to API constraints limiting queries to 0.01 square degrees per request, the study area was subdivided into a grid of smaller tiles (0.08° × 0.08°), and images were retrieved iteratively with deduplication to avoid redundant entries.

For each image, the following metadata were extracted and stored: a unique image identifier, geographic coordinates (latitude and longitude), capture timestamp, compass angle (camera orientation), and the image URL for download. Images were downloaded at a resolution of 2048 × 1536 pixels to balance detail retention with processing efficiency. The final dataset comprised 14,327 unique geotagged images distributed across the study area.

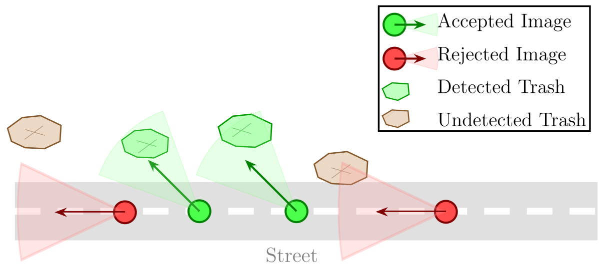


Figure 2. Schematics of possible detection outcomes using SVI to corresponding roadside waste locations.

Depending on our SVI extraction methodology, the experimental outcomes can be categorized into four primary scenarios. First, an SVI frame may fail to capture the roadside waste entirely, leaving the accumulation undetected, as illustrated in Figure 2. Second, the waste may be fully visible within the frame and successfully detected by the framework (cf. Figure 2). Third, the waste may be fully visible but remain undetected (cf. Figure 2). Fourth, the waste may be only partially visible within the image scene, due to truncation or occlusion, leading to either a successful or failed detection, (cf. Figure 2).

4. Methodology

This study proposes a fully automated end-to-end framework for detecting and mapping illegal roadside waste dumps using a large visual foundation model, as illustrated in Figure 3. The workflow comprises: (1) imagery acquisition via the Mapillary API, (2) zero-shot detection using the Segment Anything Model 3 (SAM 3) (Carion et al., 2025), (3) confidence-based quality filtering, (4) interactive geospatial visualization, and (5) export the resulting roadside waste locations in GeoJSON[1] format, including the point coordinates of the GPS location of each Mapillary image and the timestamp at which the image was captured.

4.1 Zero-Shot Detection using a Vision Foundation Model

SAM 3 is a large vision foundation model enabling strong prediction accuracy for detection and segmentation tasks to previously unseen object categories without requiring task-specific fine-tuning or labeled training data (Carion et al., 2025). This eliminates the need for annotated training data, whose creation is both costly and time-consuming, and allows the approach to be readily transferred to other cities with Mapillary coverage.

SAM 3 supports text-prompted segmentation, in which natural language descriptions of target objects are provided as input conditions to guide segmentation and detection. We instruct the model through prompt-engineering to identify waste accumulations within each image, as shown in Figure 3. For each image, SAM 3 produces a segmentation masks corresponding to regions that matched the text prompts, together with an associated confidence score representing the model's estimated likelihood that the detected region corresponds to the specified object class. We prompt SAM 3 with the text description `trash`

[1] `https://geojson.readthedocs.io/en/latest/`

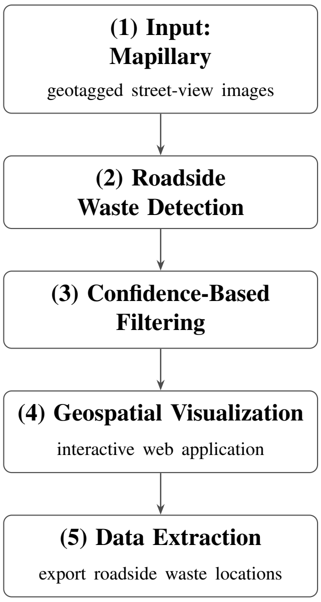


Figure 3. Overview of the proposed end-to-end framework for roadside waste detection and mapping from SVI.

`pile` to guide zero-shot segmentation of waste accumulations. For the purposes of this study, an image was classified as a successful roadside waste detection whenever the model produced a segmentation mask, regardless of the mask's spatial extent. All such detections were subsequently mapped in the interactive web-based application.

Arising from visually similar objects (e.g., construction materials, loose soil, or dense vegetation), SAM 3 can lead to false positive predictions. To reduce these a confidence threshold τ_c and mask thresholds τ_m is applied for filtering. To set τ_c and τ_m to a reasonable value that filters out enough false positives we conducted experiments testing four different confidence thresholds with the grid search algorithm.

4.2 Geospatial Visualization

For each detection with a confidence score exceeding the threshold τ_c, the geographic coordinates embedded in the image metadata are extracted and assigned as the location of the detected waste site. To visualize the spatial density of roadside waste, we employ a Kernel Density Estimation (KDE) heatmap (Silverman, 1986). Within this visualization, high-density clusters are delineated by a color gradient: darker color indicates a high counted number of waste sites within a localized area, whereas lighter color denotes regions with a lower frequency of detections by our framework. Furthermore, the spatial extent of these clusters illustrates the overall geographical footprint of the waste, highlighting areas that exhibit continuous, spatially connected waste accumulations. The resulting geospatial dataset is subsequently visualized using Folium[2], a Python library for rendering interactive maps based on the Leaflet[3] framework, as illustrated in Figure 1.

This design choice allows municipal authorities and urban planners to explore spatial heatmaps and point-layer overlays representing illegal dumping hotspots across large urban areas without specialized GIS software, as illustrated in Figure 1. The resulting interactive web-based application potentially provides stakeholders with a visual location-based decision-support tool for prioritizing cleanup operations, scheduling field inspections, and targeting enforcement interventions in areas characterized by high densities of waste accumulation.

4.3 Evaluation Dataset and Metrics

Model performance was evaluated using a publicly available waste pile dataset[4] comprising 100 annotated street-level images with instance segmentation masks in COCO format (Lin et al., 2015). Predictions were categorized as true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN), from which Precision and Recall were calculated as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}}, \qquad \text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}}. \quad (1)$$

Spatial overlap between predicted and ground-truth masks is quantified by the Intersection over Union (IoU) and the Dice coefficient:

$$\text{IoU} = \frac{|\hat{M} \cap M|}{|\hat{M} \cup M|}, \qquad \text{Dice} = \frac{2|\hat{M} \cap M|}{|\hat{M}| + |M|}, \quad (2)$$

where $\hat{M}$ and M denote the predicted and ground-truth mask regions respectively. Beside the IoU, we report the Dice coefficient as the primary overlap metric rather than mean Average Precision (mAP), which aggregates Precision-Recall curves across multiple IoU thresholds and object categories. In a single-class, zero-shot segmentation setting, mAP conflates localization and classification performance in a manner that penalizes confident detections at moderate overlap. The Dice coefficient, by contrast, directly measures the fractional overlap of predicted mask area and is more interpretable for binary segmentation tasks with irregular objects such as waste dumps, where mask boundaries are inherently imprecise.

5. Experimental Results

As illustrated in Figure 1, high-density waste clusters within our Dhaka study region are predominantly concentrated in peri-urban areas, particularly Tongi, Purbachal, and Keraniganj. Conversely, central Dhaka exhibits comparatively fewer detections, likely reflecting more comprehensive formal waste collection coverage or less dense street-view imagery coverage. Annotated examples confirm a heterogeneous composition of waste: ad-hoc roadside piles, accumulated bagged waste in commercial corridors and riverbank deposits that pose direct water contamination risks, as demonstrated by the examples available in our web application.

5.1 Quantitative Performance

To maximize the detection of roadside waste while simultaneously limiting the number of false positives, we employed a grid-search strategy for hyperparameter tuning to optimize the F1-score. Specifically, we evaluated SAM 3 on the previously described Roboflow dataset for all combinations of confidence thresholds $\tau_c \in \{0.1, 0.2, \ldots, 0.9\}$ and mask thresholds

[2] `https://python-visualization.github.io/folium/latest/`

[3] `https://leafletjs.com/`

[4] `https://universe.roboflow.com/sgugit-vl4tb/trash-mezg5`

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XII-4-W2-2026-9-2026 | © Author(s) 2026. CC BY 4.0 License.

$\tau_m \in \{0.1, 0.2, \ldots, 1.0\}$. The best-performing configuration, measured by the highest F1-score, is reported in Table 1. This configuration corresponds to $\tau_c = 0.7$ and $\tau_m = 0.1$, yielding an F1-score of 0.54.

Table 1. Highest detection rate of SAM 3 achieved after hyper-parameter fine tuning with a grid search strategy.

τ_c	τ_m	Acc.	Pre.	Rec.	F1	mIoU	mDice
0.7	0.1	92.0%	0.551	0.533	0.542	0.820	0.894

The effect of the mask threshold is visible qualitatively in Figure 4. At $\tau_c = 0.7$ and $\tau_m = 0.1$, only mask regions with the highest predicted confidence are retained, producing compact predictions that closely follow object boundaries. As shown, the prediction closely follows the ground truth boundary.

(a) ground truth annotation (b) SAM 3 prediction

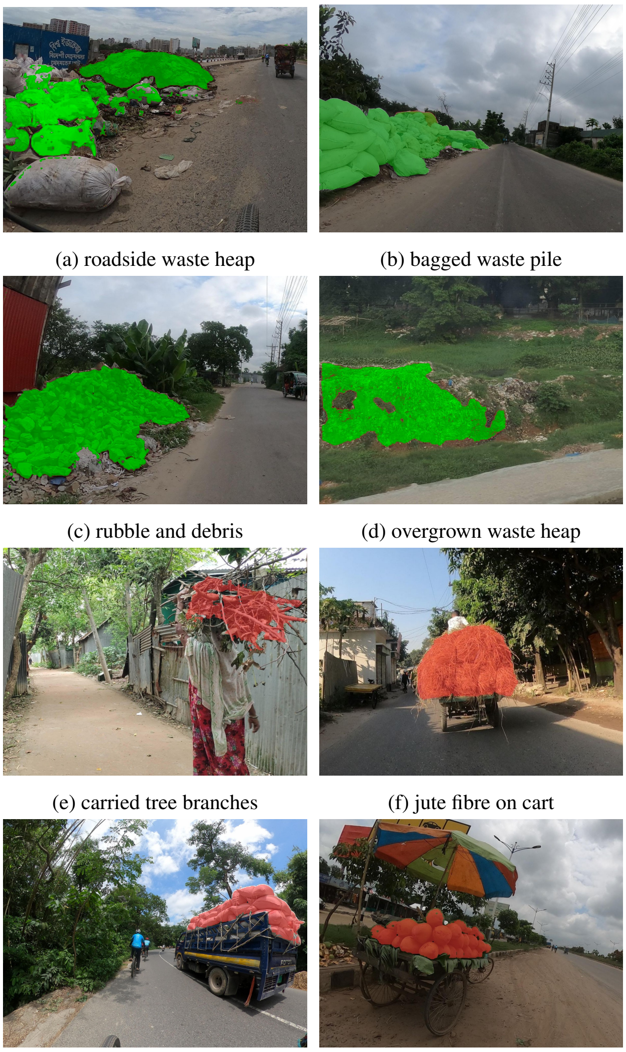
Figure 4. Comparison of ground truth annotation and SAM 3 zero-shot segmentation on a sample image from the Roboflow dataset at $\tau_c = 0.7$ and $\tau_m = 0.1$.

5.2 Qualitative Analysis

The Mapillary street-level imagery in Dhaka encompasses a broad range of visual conditions, from clearly delineated roadside waste accumulations to highly cluttered urban scenes in which waste boundaries are ambiguous. The following examples illustrate the predictions of SAM 3 across this diverse spectrum, thereby highlighting both the generalization capability and the limitations of zero-shot segmentation in this context.

5.2.1 True Positive Detections SAM 3 demonstrated notable generalization across waste morphologies without any task-specific training. In Figure 5a, the model accurately delineates a large mixed organic and inorganic waste heap deposited on a roadside verge, correctly excluding the adjacent road surface and vehicles. In another example, Figure 5b shows precise segmentation of densely packed polyethylene waste bags stacked against a wall, where the model successfully resolved individual bag boundaries. In Figure 5c, construction rubble and demolition debris accumulated on a peri-urban lane are correctly identified despite their irregular geometry and absence of conventional waste-like appearance, suggesting the model responds to contextual cues beyond color alone. The example in Figure 5d is particularly noteworthy: An open dump located at a distance from the road, almost entirely overgrown by vegetation and adjacent to a drainage channel, is successfully masked, demonstrating that the model is not limited to proximity and can detect spatially displaced dumps visible in the image periphery.

5.2.2 False Positive Detections In the example illustrated in Figure 5e, tree branches being carried on a person's head produce a loose, irregular silhouette that the model misinterprets as organic waste. Figure 5f illustrates a wrong segmentation of bulk fibrous materials, i.e., jute fibre on a transport cart,



(a) roadside waste heap (b) bagged waste pile
(c) rubble and debris (d) overgrown waste heap
(e) carried tree branches (f) jute fibre on cart
(g) bundled sacks on cargo truck (h) fruit vendor cart

Figure 5. True and False Positive segmentation results on exemplary Mapillary street-view imagery from Dhaka.

which resembles bundled waste in texture and morphology. In Figure 5g a truck transporting large bundled sacks of soft materials are incorrectly detected as waste, presumably due to their compact, clustered mass and texture closely resembling bagged waste piles. Finally, Figure 5h shows a roadside fruit vendor's cart triggering a false segmentation due to the compact mass.

A key limitation of the proposed approach lies in its lack of scene-level contextual understanding, which leads to false positives for visually similar objects such as construction materials, vegetation, or objects carried by humans. Future work could integrate scene understanding to incorporate contextual constraints, such as distinguishing between static waste accumulations and objects associated with human activity or transportation. For instance, detections associated with moving objects or human carriers could be filtered using temporal or semantic cues.

5.3 Precise Roadside Waste Extent Localization

In this work, we used only the GPS coordinates associated with each image as input for the heatmap generation. However, mapping detected sidewalk waste could benefit substantially from going beyond simple localization and instead identifying not only the presence of waste in an image, but also its precise spatial extent and quantity. Such information could further improve and optimize waste management strategies.

At the same time, estimating the spatial dimensions of waste from images is inherently challenging. This difficulty arises from factors such as occlusions, varying illumination, and changing weather conditions, among others. Although the proposed approach can estimate the spatial extent of detected roadside waste, accurate placement requires that this extent, for example in the form of a polygon, be mapped to the correct location in 2D space. Achieving this requires depth estimation.

Depth estimation is particularly challenging when image data are acquired with heterogeneous and uncalibrated camera systems, for which the intrinsic and extrinsic parameters are unknown, as is the case for Mapillary SVI. In addition, image quality varied substantially due to different acquisition platforms, including bicycle-, pedestrian-, and vehicle-based recordings. As conventional depth estimation methods relying on calibrated camera systems and multi-view geometric reconstruction are not applicable to such heterogeneous image data, a deep learning-based approach was used to estimate the depth of segmented sidewalk waste in the images. Specifically, we employed a fine-tuned version of Depth Anything V2 (Yang et al., 2024) for outdoor metric depth estimation[5]. The model leverages a Dense Vision Transformer architecture (Ranftl et al., 2021) with a DINOv2 backbone (Oquab et al., 2023).

The resulting depth information remains uncertain and of limited accuracy, as shown in Figure 6. While no depth is estimated for the roadside waste on the left-hand side, a depth value is assigned to the roadside waste on the right-hand side. However, because the waste extends along the street, the method provides only a single depth value rather than a depth range, making it unclear which part of the accumulation this value represents. This limitation is partly caused by the uncalibrated nature of the SVI data and the resulting propagation of uncertainties in the image. Additionally, since SVI captures only the camera-facing surface rather than the full volume and spatial extent of the waste accumulation, precise area estimation is not feasible.

5.4 Application to other Urban Areas

To evaluate whether the proposed pipeline transfers to urban environments beyond the primary study area, we applied the identical zero-shot configuration ($\tau_c = 0.7$, $\tau_m = 0.1$) to Mapillary street-view imagery from Nairobi, Kenya, without any retraining or threshold adjustment. Of 7,228 processed images, 350 (4.8%) were flagged as containing roadside waste, demonstrating that the framework remains operational and produces candidate detections in a substantially different urban and geographic context without modification.

To assess the reliability of these detections, we conducted a manual review of a random sample of 100 flagged images. Of these, 32 were confirmed as genuine roadside waste, while the majority reflected confusion with locally common visually similar objects, consistent with the false-positive patterns observed in Dhaka (Sec. 5.2.2). These results indicate that the zero-shot approach transfers to a new city at the level of producing plausible candidate detections without any retraining.

[5] `https://huggingface.co/depth-anything/Depth-Anything-V2-Metric-Outdoor-Large-hf`

Figure 6. Representative Mapillary image captured from an in-vehicle perspective, along with the corresponding roadside waste segmentation and successfully predicted depth with respect to the camera's GPS coordinates.

(a) roadside waste heap (b) roadside sand pile

Figure 7. Representative detections from Nairobi, Kenya

6. Conclusion & Outlook

In this paper, we present an end-to-end framework for the detection and mapping of roadside waste in rapidly expanding urban areas. Utilizing street-view imagery (SVI) from Mapillary, our approach leverages SAM 3 for zero-shot semantic segmentation. The proposed framework culminates in an interactive web application that visualizes the spatial density of roadside waste, providing a valuable location-based decision-making system for optimizing and monitoring municipal waste management.

However, the Mapillary street-view coverage in Dhaka is incomplete, especially in narrow, frequently inaccessible alleys and the interiors of informal settlements, despite these areas potentially exhibiting a higher prevalence of waste accumulation. Moreover, the heterogeneity in image capture dates between 2023 and 2025 implies that the resulting detections may represent historical rather than contemporaneous conditions. However, the proposed approach is equally transferable to more up-to-date SVI data sources, where available. Future improvements in the availability of up-to-date street-view imagery could be achieved by complementing crowdsourced campaigns with organized, systematic data acquisition, for example, using cameras mounted on government or other authoritative vehicles to provide more consistent and comprehensive spatial coverage. Our analysis reveals that while zero-shot segmentation is primarily constrained by a lack of domain-specific contextual reasoning, the scarcity of annotated segmentation masks for roadside waste renders effective fine-tuning highly challenging. Consequently, zero-shot segmentation remains the most pragmatic choice for scalable deployment.

References

Arzoumanidis, L., A. Samee, A. M., Kanna, E., Nguyen, S. H., Dehbi, Y., 2026. Domain-Adaptive Object Detection for Enriching Semantic 3D City Models with Building Storeys from Street-View Images. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV/1/WG-2026.

Bahauddin, K., Uddin, M., 2012. Prospect of Solid Waste Situation and an Approach of Environmental Management Measure (EMM) Model for Sustainable Solid Waste Management: Case Study of Dhaka City. *Journal of Environmental Science and Natural Resources*, 5(1), 99–111.

Biljecki, F., Ito, K., 2021. Street View Imagery in Urban Analytics and GIS: A Review. *Landscape and Urban Planning*, 215, 104217.

Carion, N., Gustafson, L., Hu, Y.-T., Debnath, S., Hu, R., Suris, D., Ryali, C., Alwala, K. V., Khedr, H., Huang, A., Lei, J., Ma, T., Guo, B., Kalla, A., Marks, M., Greer, J., Wang, M., Sun, P., Rädle, R., Afouras, T., Mavroudi, E., Xu, K., Wu, T.-H., Zhou, Y., Momeni, L., Hazra, R., Ding, S., Vaze, S., Porcher, F., Li, F., Li, S., Kamath, A., Cheng, H. K., Dollár, P., Ravi, N., Saenko, K., Zhang, P., Feichtenhofer, C., 2025. Sam 3: Segment Anything with Concepts.

Cicala, L., Gargiulo, F., Parrilli, S., Amitrano, D., Pigliasco, G., 2024. Progressive Monitoring of Micro-Dumps Using Remote Sensing: An Applicative Framework for Illegal Waste Management. *Sustainability*, 16(13), 5695.

Dong, S., Xu, W., Zhang, H., Gong, L., 2025. Cot-DCN-YOLO: Self-Attention-Enhancing YOLOv8s for Detecting Garbage Bins in Urban Street View Images. *The Egyptian Journal of Remote Sensing and Space Sciences*, 28(1), 89-98.

Du, L., Zuo, J., Vanzo, J., Chang, R., Zillante, G., 2023. Assessing and predicting the illegal dumping risks in relation to road characteristics. *Waste Management*, 169, 332-341.

Fan, Z., Feng, C.-C., Biljecki, F., 2025. Coverage and Bias of Street View Imagery in Mapping the Urban Environment. *Computers, Environment and Urban Systems*, 117, 102253.

Fazzo, L., De Santis, M., Beccaloni, E., Scaini, F., Iavarone, I., Comba, P., Airoma, D., 2020. A Geographic Information System-Based Indicator of Waste Risk to Investigate the Health Impact of Landfills and Uncontrolled Dumping Sites. *International journal of environmental research and public health*, 17(16), 5789.

Hossain, M. Y., Nijhum, I. R., Sadi, A. A., Shad, M. T. M., Rahman, R. M., 2021. Visual Pollution Detection Using Google Street View and YOLO. *2021 IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, 0433–0440.

Hou, Y., Biljecki, F., 2022. A Comprehensive Framework for Evaluating the Quality of Street View Imagery. *International Journal of Applied Earth Observation and Geoinformation*, 115, 103094.

Ichipi, E. B., Senekane, M. F., 2023. An Evaluation of the Impact of Illegal Dumping of Solid Waste on Public Health in Nigeria: A Case Study of Lagos State. *International Journal of Environmental Research and Public Health*, 20(22).

Kim, Y., Cho, J., 2022. AIDM-Strat: Augmented Illegal Dumping Monitoring Strategy through Deep Neural Network-Based Spatial Separation Attention of Garbage. *Sensors*, 22(22), 8819.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., Dollár, P., Girshick, R., 2023. Segment Anything.

Li, W., Dai, J., 2026. DDCN-YOLO: A Multi-Dimensional Detection Model for Garbage Bins based on Relation Enhancement and Dynamic Perception. *Measurement*, 257, 118659.

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S., 2017. Feature Pyramid Networks for Object Detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2117–2125.

Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L., Dollár, P., 2015. Microsoft COCO: Common Objects in Context.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B., 2021. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows.

Ma, R., Yang, C., Chen, J., Biljecki, F., Li, X., 2026. Street View Imagery-Based Method for Reconstructing 3D Building Façade Openings. *Automation in Construction*, 184, 106842.

Moral, P., Álvaro García-Martín, Escudero-Viñolo, M., Martínez, J. M., Bescós, J., Peñuela, J., Martínez, J. C., Alvis, G., 2022. Towards Automatic Waste Containers Management in Cities via Computer Vision: Containers Localization and Geo-Positioning in City Maps. *Waste Management*, 152, 59-68.

Ngalo, N., Thondhlana, G., 2023. Illegal Solid-Waste Dumping in a Low-Income Neighbourhood in South Africa: Prevalence and Perceptions. *International Journal of Environmental Research and Public Health*, 20(18).

Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., Assran, M., Ballas, N., Galuba, W., Howes, R., Huang, P.-Y., Li, S.-W., Misra, I., Rabbat, M., Sharma, V., Synnaeve, G., Xu, H., Jegou, H., Mairal, J., Labatut, P., Joulin, A., Bojanowski, P., 2023. DINOv2: Learning Robust Visual Features without Supervision. *arXiv preprint arXiv:2304.07193*.

Osco, L., Wu, Q., de Lemos, E. L., Gonçalves, W., Ramos, A. P., Li, J., Junior, J. M., 2023. The Segment Anything Model (SAM) for Remote Sensing Applications: From Zero to One Shot. *ArXiv*, abs/2306.16623.

Paulo, D. J., Martins, J. a., Proença, H., Neves, J. a. C., 2026. Streetview-waste: A multi-task dataset for urban waste management. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 3015–3025.

Ranftl, R., Bochkovskiy, A., Koltun, V., 2021. Vision transformers for dense prediction. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 12159–12168.

Silverman, B. W., 1986. *Density estimation for statistics and data analysis*. Chapman & Hall/CRC monographs on statistics and applied probability, Chapman and Hall, London.

Sun, X., Yin, D., Qin, F., Yu, H., Lu, W., Yao, F., He, Q., Huang, X., Yan, Z., Wang, P., Deng, C., Liu, N., Yang, Y., Liang, W., Wang, R., Wang, C., Yokoya, N., Hänsch, R., Fu, K., 2023. Revealing Influencing Factors on Global Waste Distribution via Deep-Learning based Dumpsite Detection from Satellite Imagery. *Nature Communications*, 14.

Torres, R. N., Fraternali, P., 2021. Learning to Identify Illegal Landfills through Scene Classification in Aerial Images. *Remote. Sens.*, 13, 4520.

Wikurendra, E. A., Csonka, A., Nagy, I., Nurika, G., 2024. Urbanization and Benefit of Integration Circular Economy into Waste Management in Indonesia: A Review. *Circular Economy and Sustainability*, 4(2), 1219-1248.

Yang, B., Lu, W., Chen, J., Yuan, L., Bao, Z., 2025. Dynamics of Street Views and Socio-Economic Conditions in Profiling Illegal Dumping 'Black Spots': An LLM-Enabled Study in Hong Kong. *Waste Management*, 207, 115137.

Yang, L., Kang, B., Huang, Z., Xu, X., Feng, J., Zhao, H., 2024. Depth Anything: Unleashing the Power of Large-Scale Unlabeled Data. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 10371–10381.

Yong, Q., Wu, H., Wang, J., Chen, R., Yu, B., Zuo, J., Du, L., 2023. Automatic Identification of Illegal Construction and Demolition Waste Landfills: A Computer Vision Approach. *Waste management*, 172, 267-277.

Zhang, S., Ma, J., 2026. Urban Waste Management Inequities in New York City: A Case Study on Spatial Patterns of Controlled and Uncontrolled Street Waste and their Associated Factors. *Waste Management*, 212, 115333.

Zhang, S., Ma, J., Jiang, F., 2025. Monitoring Street-Level Improper Dumpsites via a Multi-Modal and LLM-Based Framework. *Resources, Conservation and Recycling*, 218, 108227.

Zhang, Z. et al., 2024. Municipal Solid Waste Management Challenges in Developing Regions: A Comprehensive Review and Future Perspectives for Asia and Africa. *Science of the Total Environment*, 930, 172794.

Zheng, X., Amemiya, M., 2023. Method for Applying Crowdsourced Street-Level Imagery Data to Evaluate Street-Level Greenness. *ISPRS International Journal of Geo-Information*, 12(3), 108.