UrbanWaste: In-the-Bin Dataset for Waste Disposal Inspection with Multi-Granularity Hierarchical Labels

Zhuoqi Ma[1,2], Zejun You[1,2], Yang Dong[1,2], Yukai Liu[2], Xiyue Gao[2], Qiguang Miao[1,2*]

[1]Xi'an Key Laboratory of Big Data and Intelligent Vision
[2]School of Computer Science and Technology
Xidian University
Xi'an, Shaanxi, 710126, China
zhuoqima@xidian.edu.cn, qgmiao@xidian.edu.cn

Abstract

Our world faces the challenge of efficiently and responsibly managing the ever-growing volume of urban waste. Many countries and regions have implemented categorized trash bins and require residents to sort their waste according to specified criteria. Proper waste classification by residents significantly reduces the workload in the waste disposal process. However, due to the lack of effective supervision during classification, the quality of waste sorting is often compromised. This misclassification can lead to higher pollution risks, lower recycling rates, and increased waste management costs and difficulties. To address this issue, we propose using images captured from within trash bins to supervise garbage delivery. We introduce ***UrbanWaste***, an image dataset specifically designed for in-the-bin waste detection and segmentation. The dataset includes 25,254 RGB images and 140,008 annotated items, featuring dense annotations and multi-granularity labels across 193 distinct waste categories. We evaluated state-of-the-art segmentation models to understand their generalization and performance on UrbanWaste. Based on this dataset, we developed a comprehensive workflow for waste classification inspection, which has been deployed in real-world districts to assess the system's effectiveness. We hope ***UrbanWaste*** will inspire new directions in AI research for environmental sustainability.

Dataset — https://github.com/zma029/UrbanWaste

Introduction

As urbanization and population growth accelerate globally, the surge in urban waste is putting significant pressure on urban infrastructure and living environments. Governments and organizations are compelled to adopt more robust measures, such as waste sorting systems and recycling initiatives. However, improper waste management reduces recycling efficiency and increases costs of downstream waste processing. Faced with the escalating waste crisis, devising effective waste sorting and management strategies to mitigate environmental impacts is an urgent priority.

In the waste management pipeline, the first step is the sorting of residential waste. Proper household waste sorting and disposal can significantly reduce the burden of manual handling downstream and improve recycling efficiency. However, due to insufficient oversight, some residents fail to adhere to standard classification procedures. As a result, a significant portion of waste sorting still depends on manual labor, which is both inefficient and poses health risks due to extended exposure to hazardous or unsanitary materials.

This issue has already attracted widespread attention. Some cities have hired supervisors to monitor residents' waste classification. This manual supervision is not only costly, but adversely affect the health of the supervisors. Additionally, Qiu *et al.* (Qiu et al. 2022) have proposed using X-ray technology to inspect garbage bags. Although X-rays have excellent penetration capabilities, installing the equipment at every community disposal point would require a significant financial investment. Moreover, requiring residents to run their trash through an X-ray machine before disposal can be overly cumbersome.

Recent advancements in object detection and segmentation have spurred the development of waste classification algorithms (Mittal et al. 2016; Rabano et al. 2018; Ruiz et al. 2019), aimed at identifying misclassified waste objects. However, the limited data and label abundance in existing datasets have significantly constrained the practicality of these methods. Some waste classification datasets feature clean, uncluttered backgrounds (Thung and Yang 2016; Koskinopoulou et al. 2021), which poses generalization challenges for models trained solely on such data when applied to realistic waste management systems. Furthermore, several in-the-wild datasets primarily focus on the classification of a few recyclable waste categories (Bashkirova et al. 2022; Proença, Quintas, and Dias 2020; Koskinopoulou et al. 2021), thus failing to encompass the diverse range of urban waste categories.

These observations inspired us to propose *UrbanWaste*, an image dataset focusing on in-the-bin waste detection and segmentation, as illustrated in Figure 1. *UrbanWaste* comprises 25,254 images taken from inside garbage bins in actual residential areas. Based on the household waste classification guidelines in Shanghai, China (Zhou et al. 2019), we establish a hierarchical labeling system and annotated solid objects, with finally an average of 5.54 labeled instances per image. *UrbanWaste* is designed to mirror the real-world di-

*Corresponding Author.

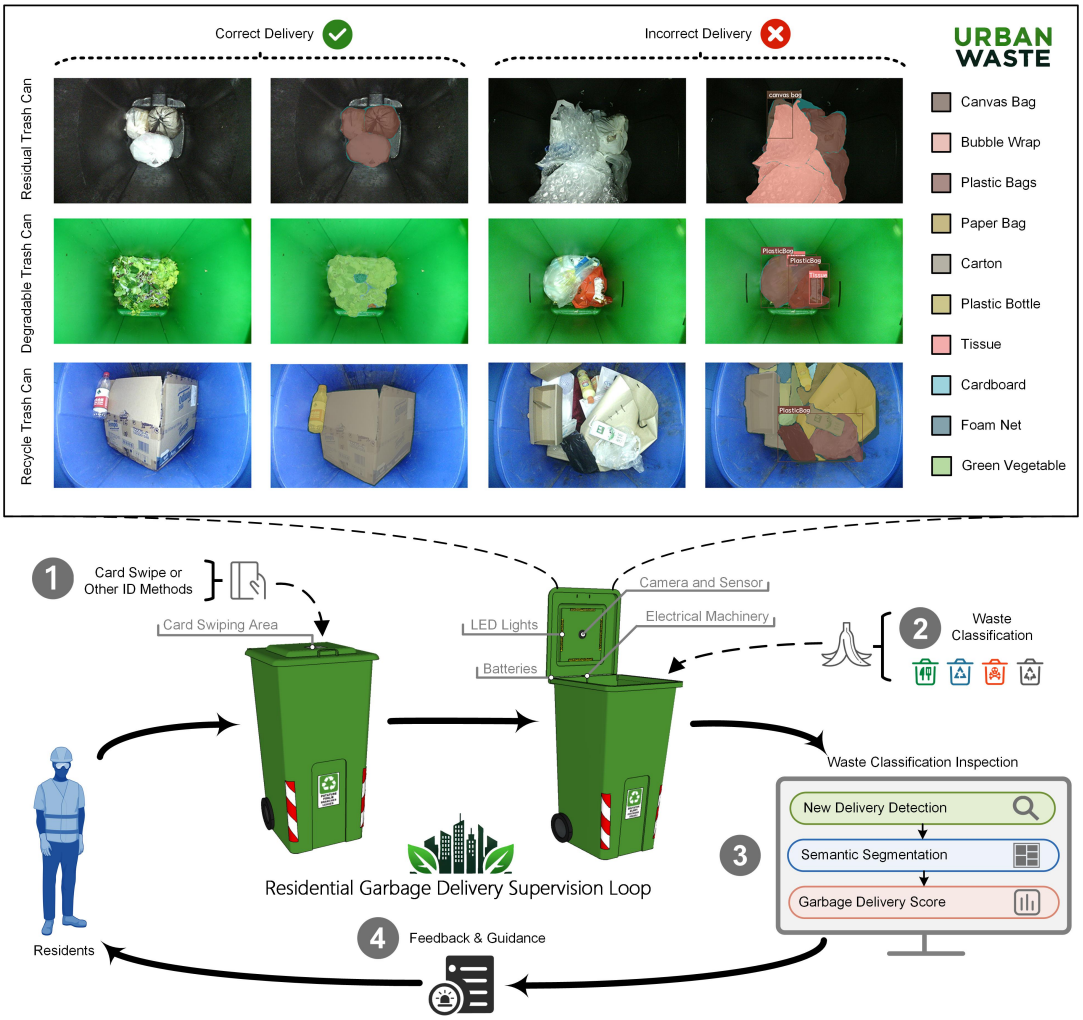

Figure 1: Illustration of UrbanWaste dataset and garbage delivery supervision process.

versity and complexity of waste classification, ensuring it captures the wide range of scenes and types of waste encountered in everyday environments.

In addition, we propose a novel method for garbage delivery supervision. As shown in Figure 1, when residents deliver their garbage, the camera in the garbage bins would took the photo of current delivery. Then the waste classification inspection module would give score based on change detection and semantic segmentation. This feedback help residents better understand the importance and correct methods of waste sorting. In summary, our contributions are:

1. We introduce the UrbanWaste dataset, which fills the data gap in addressing the challenge of detecting waste within garbage bins. UrbanWaste features dense annotations and multi-granularity labels covering a wide range of waste categories. It surpasses existing waste detection datasets in scale, label diversity, and background complexity.
2. We propose a novel waste classification inspection method which combines change detection with semantic segmentation model to effectively address the issue of automated supervision of residents' garbage delivery process.
3. Experimental results demonstrate that the UrbanWaste dataset effectively supports the training of detection and segmentation baselines, surpassing other datasets in enhancing model performance and generalization capabilities in practical applications.
4. Waste Classification inspection model trained on the UrbanWaste dataset have already been deployed in real residential communities.

Related Work

Waste Classification Datasets

In contrast to traditional benchmarks for object detection and segmentation, real-world waste classification presents unique challenges. These include complex backgrounds, a wider range of waste categories, and issues related to object deformation and occlusion. These specific requirements necessitate datasets that are tailored to more complex and specific scenarios for waste classification.

DataSet	Scale	Category	Annotation type	Hierarchy label
TrashNet	2527	6	Classification	✘
FloW	2000	1	bbx	✘
Taco	1500	60	bbx & Mask	✘
ZeroWaste	4661	4	bbx & Mask	✘
ETHSeg	5038	12	bbx & Mask	✘
ReSortIT	16000	4	bbx & Mask	✘
Labeled Waste in the Wild	2527	3	bbx & Pixel Level	✘
UrbanWaste	25254	193	bbx & Mask	✔

Table 1: Comparison of waste image datasets.

Several datasets have been created to explore waste classification and detection. TrashNet (Thung and Yang 2016) stands out as one of the earliest publicly available datasets for waste classification. In TrashNet, images are captured against a clean white poster board background, with each image focusing on a single waste item. It is specifically curated for image-level classification and may not be ideal for tasks involving the precise localization of garbage objects. To facilitate detection tasks, Labeled Datasets in the Wild (Sousa, Proença, and Dias 2019) provides object category labels and bounding box annotations. These images were captured in various real-world environments and conditions. TACO (Proença, Quintas, and Dias 2020), on the other hand, collects images in outdoor scenes, featuring one or a few foreground waste objects with minimal occlusion. However, this characteristic makes it less practical for real-world waste classification scenarios.

Both ReSort-IT (Koskinopoulou et al. 2021) and ZeroWaste (Bashkirova et al. 2022) were designed for waste object detection on conveyor belts. In contrast to ReSort-IT, which uses synthetic backgrounds to highlight the target objects, ZeroWaste presents real conveyor belt images from waste recycling facilities. Nevertheless, both of these datasets focus solely on recyclable waste, which can not adequately represent the diversity and complexity encountered in waste management systems. ETHSeg (Qiu et al. 2022) proposed waste inspection using X-ray images to address occlusion issues by utilizing X-ray penetration. However, the expensive cost of X-ray equipment and the health risks associated with radiation for operators would contradict the original intention to create a safer and cost-effective waste classification pipeline.

From Table 1, *UrbanWaste* excels in almost every aspect compared to existing waste image datasets, including scale, annotation richness, diversity, and background authenticity.

Waste Detection and Segmentation

In the past decade, there have been remarkable advancements in the field of image detection and segmentation. For semantic segmentation, DeepLabv3+ (Chen et al. 2017) has enhanced performance in semantic segmentation by leveraging atrous convolutions and refining modules with image-level features. Another architecture, the Segmentation Transformer (Yuan, Chen, and Wang 2020) addresses context aggregation using Object Context Representation (OCR). The Swin Transformer (Liu et al. 2021) optimizes modeling and latency with shifting windows and a hierarchical structure. The influential R-CNN series (Girshick 2015; He et al. 2017; Cai and Vasconcelos 2018; Pang et al. 2019) utilizes high-capacity CNNs and region proposals for object localization and segmentation, laying the foundation for subsequent segmentation and detection tasks.

Many outstanding one-stage object detection methods have been proposed, YOLO series (Redmon et al. 2016; Redmon and Farhadi 2018, 2017; Bochkovskiy, Wang, and Liao 2020) revolutionized object detection with its real-time capabilities. RetinaNet (Lin et al. 2017) incorporates a Feature Pyramid Network (FPN) to handle objects at different scales and employs a Focal Loss to address class imbalance issues. TridentNet (Li et al. 2019) proposed a multi-granularity object detection algorithm designed to enhance detection performance, especially for small objects. TridentNet significantly boosts detection performance for small objects through multi-level feature fusion and multi-scale attention mechanisms.

Automatic garbage classification algorithms have also made significant progress. Several object detection-based methods (Mittal et al. 2016; Rabano et al. 2018; Ruiz et al. 2019; Chen et al. 2020) were proposed to identify misclassified waste items. However, due to the limitation of domain-appropriate datasets, these methods mainly focused on recyclable waste.

Construction of UrbanWaste

Data Collection

To tackle the garbage delivery supervision challenge, we propose taking images of the garbage inside bins and performing detection on garbage delivery images, thereby serving a supervisory role. In order to construct a waste classification dataset that accurately reflects the real-world scenarios, we collect images directly from the in-use trash cans of residential communities. Figure 1 gives a illustration of the data collection system process. The image is collected during residents' everyday garbage delivery. The resident first swipe his/her card to unlock the system and lift up the lid, then they throw out their garbage. After that, the lid would slowly close. After the lid is closed, the camera would take a picture of the waste in the bin and upload the collected information to server.

The lids are made removable, then we install them on the normal trash cans. As shown in Figure 1, the lid consists of card swiping area to unlock the lid, camera and sensor which capture the in-the-bin image, a electrical machinery to lift up and lowering the lid, and battery modules to power for the whole system. We had a camera installed on the inside of the bin lid. We use 5 megapixel auto focus camera with zoom lens to take the picture. The size of the camera is 9×62mm. In order to capture the situation inside the opaque garbage bin, we installed LED lights around the camera for lighting. The LED light bulb is assembled into a light strip for every three bulbs. Then we stick the light strips around the camera with tapes to reduce glare.

The data collection has been ongoing for a year, more than

27,000 images with size 600x800 were collected. The collected images were sharpened using Gaussian Pyramid to enhance details. We also removed images which is damaged with severe noise and exposures.

Multi-Granularity Hierarchical Labels

Based on Shanghai's household waste classification guidelines, we define multi-granularity hierarchical labels for UrbanWaste, structured into four levels. Figure 3 illustrates the top three levels. The coarsest level comprises four categories: degradable, recyclable, residual, and hazardous. The second level refines these by materials, such as recyclable plastics or papers, while the third focuses on shapes, like paper bags or boxes. such as paper bags or paper boxes. The finest granularity is constructed during annotation. The finest granularity aligns with annotators' habits for efficiency. For example, labeling "tofu" (4_{th} level) is more intuitive than "grain products" (3_{rd} level), which simplifies the annotation process. We find that fine-level feature betters the learning of coarse-level recognition. For example, when beans were labeled as vegetables, they were always misclassified as residual wastes. When we assign "beans" as the leaf node of "vegetables", they can be correctly classified as "degradable wastes".

By establishing multi-granularity hierarchical labels, we re-envisage the traditional setting of waste detection, transitioning from single-label instance detection to coarse-to-fine hierarchical label. So that the label becomes "residual" $\rightarrow$ "shell" $\rightarrow$ "hard fruit shell" $\rightarrow$ "coconut shell". Multi-granularity hierarchical labels offers several advantages. First, it allows for the subdivision of waste classification tasks into different levels of subtasks. This enables more precise identification and classification of various types of waste. On the other hand, multi-granularity hierarchical labels can be adjusted according to different application scenarios and requirements. For example, some applications may require more detailed waste classification, while others may only need general classification. Hierarchical labels allow the system to adapt to different classification levels. On top of that, multi-granularity hierarchical labels can better organize and manage waste classification datasets. The hierarchical structure of labels makes the dataset more structured and easier to maintain and expand. This makes building data integration a sustainable task and is helpful for building large-scale waste classification datasets.

Manual Annotation

We recruited an annotation group from local company with main business in intelligent garbage classification. The annotation group works onsite to annotate the collected images. We labeled instance segmentation masks for solid objects (in contrary to liquid or gel), and use the instance mask to generate bounding boxes.

Based on the hierarchical multi-granularity labels, The participants would draw the polygon contour of the waste objects. Then, they choose the leaf label from the label hierarchy for the objects, which will automatically contain its coarse-to-fine hierarchical label path (*e.g.* "residual" $\rightarrow$ "shell" $\rightarrow$ "hard fruit shell" $\rightarrow$ "coconut shell"). When the annotation group encountered unseen categories, we will add the new label into the original class labels. Therefore, the category labels is on dynamic maintenance through out the research.

To prevent issues such as missed or incorrect labeling, we established a uniform annotation order: For each image, first, we outline the region of garbage bin, then we annotate the cluttered entities such as leftovers or vegetable scraps, finally we annotate the contour of solid waste objects (eg. plastic bags, tissue, boxes). If an object is partially occluded by other objects, we only annotate the unoccluded regions.

The annotation is saved in JSON format with the information of object label, label id, polygon masks. All annotations are checked to ensure that no labels or polygon coordinates are missing. And each image and its annotations was cross-validated by the annotation group.

Properties of UrbanWaste

UrbanWaste features complex image backgrounds, diverse category labels, and hierarchical multi-granularity labels, offers new challenges and opportunities for the advancement of waste inspection algorithm. To our knowledge, this is the first in-the-bin waste classification inspection datasets. Since hazardous bins generated little to no images, thus we mainly focused on recyclable, degradable and residual bins. Since our images is collected from actual in-use garbage bins, the background of our dataset has extreme clutterness, which is close to the actual environments of garbage disposal stream, allowing models to be trained and tested under more realistic conditions. This helps models better adapt to real-world garbage classification tasks.

UrbanWaste consist of 25,254 images, with totally 140,008 annotated instances covering 193 fine-grained waste categories. On average, UrbanWaste contains 5.54 labeled instances per image. The dataset is divided into 80% for training, 6% for validation and 14% for testing. Table 2 summarizes part of the class-wise statistics of all splits (please refer to supplementary materials for comprehensive statistics). Images in UrbanWaste are stored in JPG format of resolution 800×600. Each image has corresponding groundtruth polygon segmentation stored in PNG format and bounding box coordinates stored in TXT format. We also provide the commonly used MS COCO (Lin et al. 2014) format for generality.

Figure 2 provides the proportion of a certain type of garbage appearing in different bins. Since UrbanWaste is collected from real communities, it well reflected the unbalanced composition of residents generated waste. Most of the recyclable and degradable garbage is delivered correctly. Still, residual garbage is often misclassified. We find that ceramic and bubblewrap got the highest probability of misclassification. Figure 2 reflects the need for effective supervision or assistance in the residential waste delivery process. We find that most wastes were thrown in degradable cans, implying that downstream automatic waste detection tasks would encounter similar situations.

The diverse set of category labels allows UrbanWaste to include multiple types of garbage, providing more comprehensive coverage of different types of waste. This enables

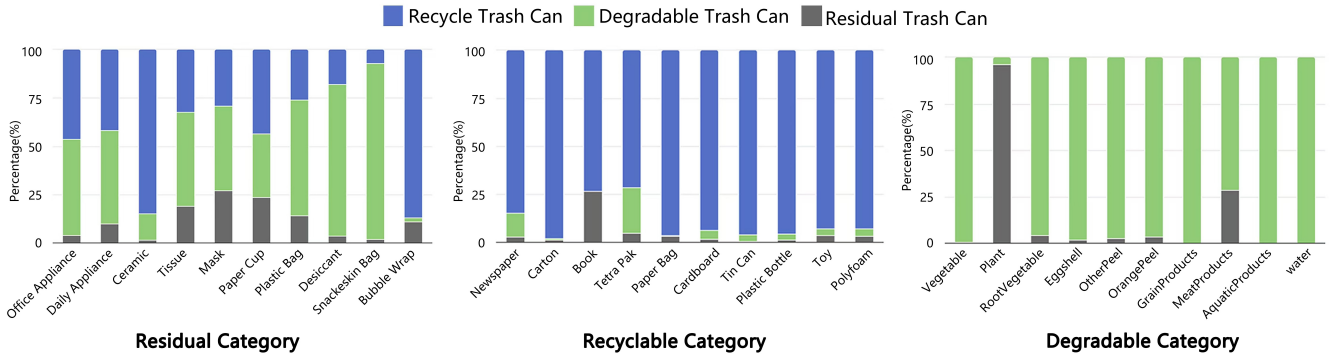

Figure 2: The proportional distribution of certain categories of garbage across different garbage bins.

	Recyclable Waste					Degradable Waste				Hazardous Waste		Residual Waste					Total
	Paper	Metal	Plastic	Glass	Fabric	Peel	Left Over	Food Waste	Plants	Drugs	Pesticides	StainedPlastic	StainedPaper	Particles	Appliance	Shell	
Train Set	8272	1018	4429	1758	626	21526	429	36088	376	183	25	25892	5278	194	1006	4204	111697
Test Set	1073	162	561	299	98	2859	67	4321	45	13	3	3202	717	42	128	492	14136
Val Set	1071	135	603	286	93	2701	54	4409	61	26	2	3292	662	58	115	548	14175
Total	20652					73268				258		45830					140,008

Table 2: Instance label statistics of the training, validation and testing splits of UrbanWaste.

models to perform more accurate classification tasks. Hierarchical multi-granularity labels provide garbage classification information at different levels of granularity, allowing models to adapt to classification tasks of varying levels. This enhances the versatility of models, making them suitable for a wide range of garbage classification applications.

UrbanWaste-Change Dataset To perform waste inspection, we collect the images taken before and after each waste delivery and construct an image-pair dataset, UrbanWaste-Change. This dataset contains 13553 pairs of images taken before and after each delivery. We annotate the newly delivered garbage with segmentation mask. The UrbanWaste-Change Dataset is annotated with UrbanWaste. After the waste object is annotated, we annotate the newly generated waste in current delivery compared with image from previous moment.

The UrbanWaste-Change dataset is specifically collected for the change detection model. Current change detection dataset (Chen and Shi 2020; Zhang et al. 2021; Wang et al. 2014) normally focus on satellite images to detect architecture or landform change. However, these methods often have unsatisfactory performance in detecting changes in daily items. The UrbanWaste-Change dataset can well support models to learn change detection in clutter backgrounds. As shown in Figure 4, by detecting the change between these image pairs and combining the results from semantic segmentation, we can determine whether current delivery has wrongly classified (eg. plastic in degradable bins).

Waste Classification Inspection

The waste classification inspection throughout the entire garbage delivery loop plays a crucial role in identifying misclassified waste objects. As shown in Figure 4, by combining segmentation results with change detection, we can decide whether the current delivery has violated waste classification regulations. The semantic segmentation module is implemented with deeplabv3+ and SAM2 (Ravi et al. 2024), and change detection is implemented with ChangeFormer(Bandara and Patel 2022).

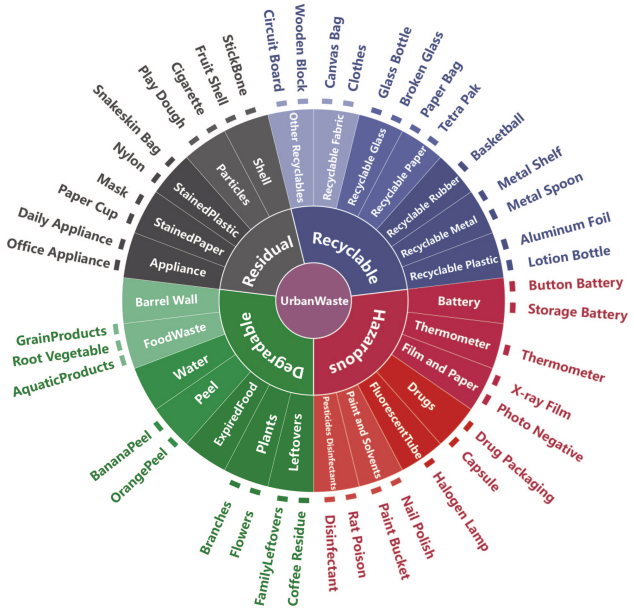

Figure 3: Illustration of multi-granularity label hierarchy

The current system has been deployed in several districts to supervise garbage delivery. Residents use their personal cards to open the garbage bin lids, and then the system will score based on the waste classification inspection result. Incorrect delivery will result in score deduction, and the system will issue warnings to residents. This feed back aims to raise residents' awareness of waste sorting.

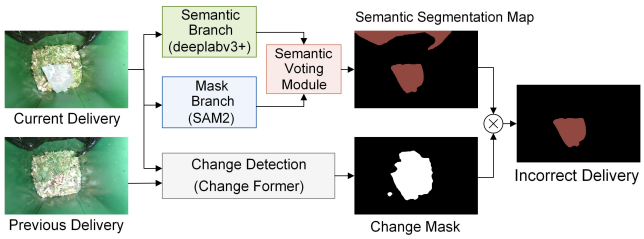

Figure 4: Illustration of waste classification inspection.

Evaluation of UrbanWaste

In this section, we provide the benchmark results on the proposed UrbanWaste dataset. Specifically, we experiment with the most widely used methods for object detection and segmentation on UrbanWaste, including TridentNet (Li et al. 2019), RetinaNet (Lin et al. 2017), Yolov5 (ultralytics 2020), Cascade R-CNN (Cai and Vasconcelos 2018), Mask R-CNN (He et al. 2017), Faster R-CNN (Girshick 2015) and DeepLabv3+ (Chen et al. 2018) for segmentation and detection. We report the experiment results in Table 3. We will provide the implementation and a detailed description of our experiments in the supplementary material.

Object Detection

Experiments For object detection experiments, we employed detectron2 toolkit (Wu et al. 2019) to implement Faster R-CNN, Mask R-CNN, Cascade R-CNN and TridentNet. We use official implementations for YOLOv5 and RetinaNet. To ensure a fair evaluation, all the compared methods use ResNet-50-FPN as the backbone. We used the pre-trained model with weights learned on MS COCO and fine-tuned it with our UrbanWaste dataset. All baseline models were fine-tuned for 300000 iterations with batch size 8 on the training set of the UrbanWaste dataset. We use the second granularity level of categories as labels for all baselines. Experiments on other granularities is provided in supplementary material. The network is trained by the Adam optimizer (Kingma and Ba 2014), where the learning rate is set as default in each baseline. All experiments are conducted on Nvidia Geforce 3090 GPU.

Results and Analysis From Table 3, it is evident that all baseline methods exhibit less-than-ideal detection accuracy, which highlights the significant challenges posed by the UrbanWaste dataset to state-of-the-art methods, with Cascade R-CNN (Cai and Vasconcelos 2018) performing slightly better than the others. In comparison to their satisfactory performance on general datasets, the cluttered backgrounds, object deformations, and occlusions are primarily responsible for the substantial degradation in performance observed in these advanced methods. This indicates that object detection algorithms still have significant room for improvement in the field of garbage detection.

Semantic Segmentation

Experiments For semantic segmentation evaluation, we experimented with DeepLabv3+, Mask R-CNN and Cascade R-CNN. For Mask R-CNN and Cascade R-CNN, we adopt the same setting as in object detection experiments. For DeepLabv3+, we used official implementation with ResNet-50 backbone and 3x3 convolutions. All baselines used ResNet-50 as the backbone and were initialized from the COCO pre-trained model. The model was finetuned for 30000 iterarions on UrbanWaste.

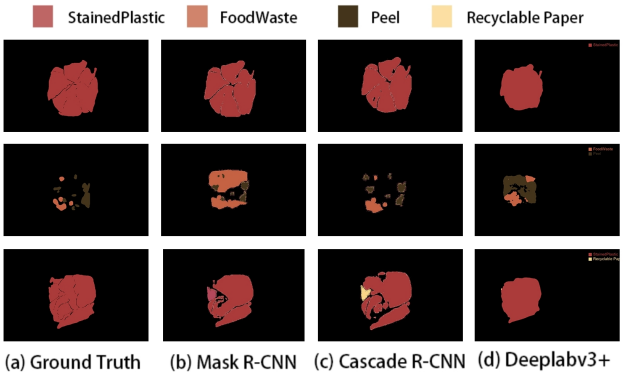

Figure 5: Qualitative comparison of segmentation on UrbanWaste. The mask color indicates the waste category.

Results and Analysis The results presented in Table 3 demonstrate that the UrbanWaste dataset also poses a challenging task for semantic segmentation. As evidenced by Table 3, DeepLabv3+ achieves relatively higher segmentation accuracy. This superior performance attests to the efficacy of dilated convolutions and multi-scale atrous spatial pyramid pooling (ASPP), which are adept at capturing multi-scale contextual information. This capability is crucial for accurately recognizing objects of varying sizes and shapes. Additionally, the encoder-decoder architecture progressively refines object boundaries using intermediate features, as illustrated in Figure 5.

Cross-Dataset Generalization Experiments

To assess UrbanWaste's generalizability, we conducted cross-dataset validation with TACO and ZeroWaste. We adopted DeepLabv3+ as our baseline model, fine-tuned it on one dataset, and subsequently validated its performance on the other datasets. Given that ZeroWaste comprises only four distinct categories (Cardboard, Soft Plastic, Rigid Plastic, Metal), we focused on classification accuracy for these specific categories when calculating statistics. Table **??** presents the mean Intersection over Union (mIOU) for the cross-validation results. Notably, the models finetuned on TACO and ZeroWaste meets a performance drop. The above results illustrate that UrbanWaste is more capable to generalize to other waste datasets. Figure 6 illustrated the segmentation results on TACO, ZeroWaste and UrbanWaste. We can observe that despite the distinct differences in backgrounds among these three datasets, models trained on UrbanWaste demonstrate strong generalizability to other datasets.

Conclusions

In this work, we introduced UrbanWaste, an image dataset focusing on the detection and segmentation of in-the-bin

Methods	Version	Detection			Segmentation				
		AP	AP_{50}	AP_{75}	AP	AP_{50}	AP_{75}	mIoU	Pixel Acc
Faster R-CNN (Girshick 2015)	ResNet-50-FPN	33.14	50.99	36.97	–	–	–	–	–
Mask R-CNN (He et al. 2017)	ResNet-50-FPN	33.00	49.60	39.19	28.97	49.10	31.99	–	–
Cascade R-CNN (Cai and Vasconcelos 2018)	ResNet-50-FPN	36.56	51.58	41.45	30.51	50.36	32.77	–	–
DeepLabv3+ (Chen et al. 2018)	ResNet-50	–	–	–	–	–	–	60.01	52.85
YOLOv5 (ultralytics 2020)	CSPDarkNet53	34.23	51.01	37.87	–	–	–	–	–
RetinaNet (Lin et al. 2017)	ResNet-50-FPN	22.49	36.25	23.32	–	–	–	–	–
TridentNet (Li et al. 2019)	ResNet-50-C4	27.75	46.18	29.74	–	–	–	–	–
Yolov11 (Jocher and Qiu 2024)	Yolov11n	–	66.81	–	–	44.62	–	–	–

Table 3: Instance segmentation and detection results on UrbanWaste.

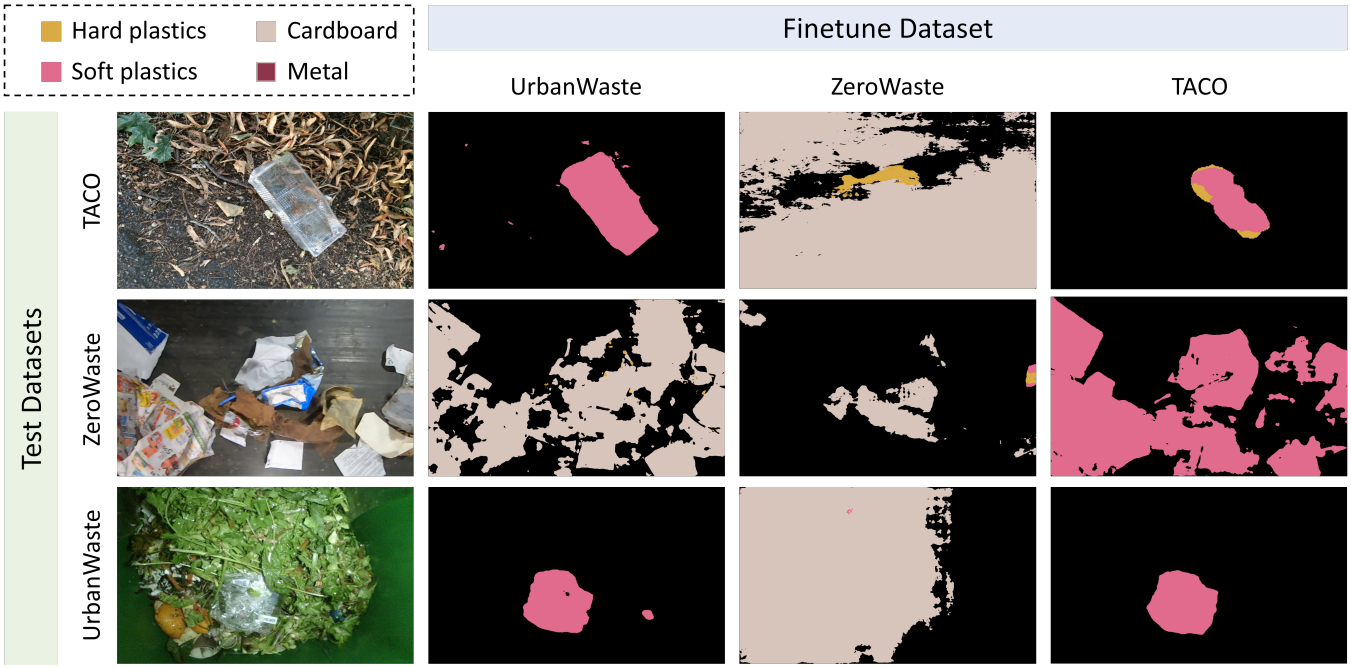

Figure 6: Cross validation of segmentation on UrbanWaste, TACO and ZeroWaste. The mask color indicates the waste category.

pretrain \ finetune	UrbanWaste	TACO	ZeroWaste
UrbanWaste	88.04	91.54 ↑	90.31 ↑
TACO	13.26 ↓	48.04	31.74 ↓
ZeroWaste	12.53 ↓	12.07 ↓	46.97

Table 4: Cross-dataset generalization between TACO and UrbanWaste, where results are repoted in mean IOU(%).

waste. UrbanWaste features real-world backgrounds and diverse category labels, presenting both challenges and opportunities for advancing garbage classification algorithms. Our experimental results show that UrbanWaste, with its rich and diverse set of samples, enhances the generalization ability of models. Based on UrbanWaste, we also propose a new workflow for automatic garbage delivery supervision.

Due to the specific characteristics of residents' lifestyles, waste distribution is inherently imbalanced. This imbalanced label distribution may lead to unfair classification and recognition of certain garbage categories, as models have limited exposure to these categories during training. Additionally, exploring the inherent coarse-to-fine hierarchical relationships among labels can potentially enhance the learning of fine-grained features, thereby improving the accuracy of the waste detection task.

In the future, we aim to expand UrbanWaste further and collect feedback from residents to assess its contributions to sustainability goals. We hope that UrbanWaste will continue to improve the precision and reliability of waste processing technologies.

Acknowledgements

This research was supported by the NSFC Foundation under Grant No. 62202360, 62302370, and the Fundamental Research Funds for the Central Universities (No. ZYTS24090).

References

Bandara, W. G. C.; and Patel, V. M. 2022. A Transformer-Based Siamese Network for Change Detection. In *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, 207–210.

Bashkirova, D.; Abdelfattah, M.; Zhu, Z.; Akl, J.; Alladkani, F.; Hu, P.; Ablavsky, V.; Calli, B.; Bargal, S. A.; and Saenko, K. 2022. ZeroWaste Dataset: Towards Deformable Object Segmentation in Cluttered Scenes. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Bochkovskiy, A.; Wang, C.-Y.; and Liao, H.-Y. M. 2020. Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*.

Cai, Z.; and Vasconcelos, N. 2018. Cascade r-cnn: Delving into high quality object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 6154–6162.

Chen, H.; and Shi, Z. 2020. A spatial-temporal attention-based method and a new dataset for remote sensing image change detection. *Remote Sensing*, 12(10): 1662.

Chen, L.-C.; Papandreou, G.; Schroff, F.; and Adam, H. 2017. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*.

Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; and Adam, H. 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, 801–818.

Chen, X.; Li, J.; Zhang, Y.; Lu, Y.; and Liu, S. 2020. Automatic feature extraction in x-ray image based on deep learning approach for determination of bone age. *Future Generation Computer Systems*, 110: 795–801.

Girshick, R. 2015. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, 1440–1448.

He, K.; Gkioxari, G.; Dollár, P.; and Girshick, R. 2017. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, 2961–2969.

Jocher, G.; and Qiu, J. 2024. Ultralytics YOLO11.

Kingma, D. P.; and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Koskinopoulou, M.; Raptopoulos, F.; Papadopoulos, G.; Mavrakis, N.; and Maniadakis, M. 2021. Robotic Waste Sorting Technology: Toward a Vision-Based Categorization System for the Industrial Robotic Separation of Recyclable Waste. *IEEE Robotics & Automation Magazine*, 28(2): 50–60.

Li, Y.; Chen, Y.; Wang, N.; and Zhang, Z. 2019. Scale-aware trident networks for object detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, 6054–6063.

Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; and Dollár, P. 2017. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, 2980–2988.

Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; and Zitnick, C. L. 2014. Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, 740–755. Springer.

Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; and Guo, B. 2021. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, 10012–10022.

Mittal, G.; Yagnik, K. B.; Garg, M.; and Krishnan, N. C. 2016. SpotGarbage: smartphone app to detect garbage using deep learning. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 940–945. ACM.

Pang, J.; Chen, K.; Shi, J.; Feng, H.; Ouyang, W.; and Lin, D. 2019. Libra r-cnn: Towards balanced learning for object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 821–830.

Proença, P. F.; Quintas, J.; and Dias, M. S. 2020. TACO: Trash Annotations in Context for Litter Detection. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 10568–10574. IEEE.

Qiu, L.; Xiong, Z.; Wang, X.; Liu, K.; Li, Y.; Chen, G.; Han, X.; and Cui, S. 2022. ETHSeg: An Amodel Instance Segmentation Network and a Real-world Dataset for X-Ray Waste Inspection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Rabano, S. L.; Cabatuan, M. K.; Sybingco, E.; Dadios, E. P.; and Calilung, E. J. 2018. Common Garbage Classification Using MobileNet. In *2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology,Communication and Control, Environment and Management (HNICEM)*, 1–4. IEEE.

Ravi, N.; Gabeur, V.; Hu, Y.-T.; Hu, R.; Ryali, C.; Ma, T.; Khedr, H.; Rädle, R.; Rolland, C.; Gustafson, L.; Mintun, E.; Pan, J.; Alwala, K. V.; Carion, N.; Wu, C.-Y.; Girshick, R.; Dollár, P.; and Feichtenhofer, C. 2024. SAM 2: Segment Anything in Images and Videos. *arXiv preprint arXiv:2408.00714*.

Redmon, J.; Divvala, S.; Girshick, R.; and Farhadi, A. 2016. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 779–788.

Redmon, J.; and Farhadi, A. 2017. YOLO9000: better, faster, stronger. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 7263–7271.

Redmon, J.; and Farhadi, A. 2018. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.

Ruiz, V.; Sanchez, A.; Velez, J. F.; and Raducanu, B. 2019. Automatic image-based waste classification. In *International Work-Conference on the Interplay Between Natural and Artificial Computation*, 422–431. Springer.

Sousa, J.; Proença, P. F.; and Dias, M. S. 2019. Automation of Waste Sorting with Deep Learning. In *2019 XV Workshop de Vis ao Computacional (WVC)*, 1–6. IEEE.

Thung, G.; and Yang, M. 2016. Classification of Trash for Recyclability Status. *Stanford University*.

ultralytics. 2020. yolov5.

Wang, Y.; Jodoin, P.-M.; Porikli, F.; Konrad, J.; Benezeth, Y.; and Ishwar, P. 2014. CDnet 2014: An expanded change detection benchmark dataset. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 387–394.

Wu, Y.; Kirillov, A.; Massa, F.; Lo, W.-Y.; and Girshick, R. 2019. Detectron2. https://github.com/facebookresearch/detectron2.

Yuan, Y.; Chen, X.; and Wang, J. 2020. Object-contextual representations for semantic segmentation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VI 16*, 173–190. Springer.

Zhang, T.; Zhang, X.; Li, J.; Xu, X.; Wang, B.; Zhan, X.; Xu, Y.; Ke, X.; Zeng, T.; Su, H.; et al. 2021. SAR ship detection dataset (SSDD): Official release and comprehensive data analysis. *Remote Sensing*, 13(18): 3690.

Zhou, M.-H.; Shen, S.-L.; Xu, Y.-S.; and Zhou, A.-N. 2019. New policy and implementation of municipal solid waste classification in Shanghai, China. *International journal of environmental research and public health*, 16(17): 3099.