RESEARCH | Article in Press

Towards automatic household food waste measurement with the smart compost bin

Aidan Beery · Patrick J. Donnelly

Received: 30 May 2026 / Accepted: 16 September 2026
© The Author(s) 2026

Abstract
Rising amounts of global food waste constitute an environmental, social, and economic disaster. Over 60% of food waste is generated by household consumers, yet we currently lack reliable means to quantify this waste. This lack of a robust household food waste measurement methodology has hindered global efforts towards meeting United Nations Sustainable Development Goal 12.3, which calls on nations to reduce global food waste by 50% by 2030. The paucity of publicly available commingled food waste image datasets has prohibited researchers from leveraging advances in computer vision for the task of automatic food waste measurement. We present an AI-enabled compost bin that measures kitchen compost waste by collecting waste images alongside environmental and weight data. We streamline the food waste annotation routine, combining weighing and imaging into a single step while enabling hands-free verbal waste labeling at time of measurement via automatic speech recognition. We provide a companion smartphone application which both provides waste analytics to users and serves as a platform for collecting instance-level food waste image annotations. To motivate a forthcoming field study—in which we will deploy 50 smart compost bins to curate a large, novel dataset of commingled food waste images—we present a computer vision model for food image recognition and demonstrate the inability of such an approach to transfer to the food waste domain.

Keywords Food waste measurement · Computer vision · Food object detection · Food waste management

1 Introduction

Food waste is a significant driver of environmental and socioeconomic harms. Estimates suggest that nearly 1.05 billion tons of edible food are lost each year, totaling $1 trillion USD in economic losses and contributing 8–10% of total greenhouse gas emissions.[1] 19% of all edible food available to consumers is wasted, with 60% of this waste occurring in households [1]. This gross inefficiency perpetuates global food insecurity, as 783 million people suffer from chronic hunger.

[1] Including both agricultural food loss and consumer food waste.

This Article in Press is shared early to give you faster access to new research. It is citable and carries a permanent DOI. The final edited version will replace it automatically.

Published online: 03 October 2026

United Nations Sustainable Development Goal (SDG) 12.3 outlines the urgent need for international collaboration to address these systemic harms generated by food waste, with the goal of reducing total volumes of food waste by 50% by 2030. However, a lack of descriptive, accurate measurements for food waste has hampered efforts to measure progress towards SDG 12.3. Policymakers and stakeholders lack information on the quantity and composition of consumer food waste streams, and this information deficit complicates the evaluation of waste reduction measures [2, 3]. The decentralized nature of household food waste streams poses a unique challenge for waste measurement studies. Existing methods for household food waste measurement typically rely on volunteers or researchers to directly weigh and categorize waste [4]. These manual measures are expensive and time-consuming, with low inter-study correlation due to poor standardization in waste measurement protocols [5].

To address the scarcity of annotated food waste data, we design and manufacture a smart compost bin to automatically measure household food waste. This consumer-scale compost bin is designed to accurately quantify food waste from individual households, providing detailed information about waste behaviors. Using a combination of imaging and environmental sensors, this device scans each disposal, providing stereoscopic image pairs alongside thermal images, ambient temperature and humidity, total volatile organic compound (tVOC) count, and disposal weight. We design a two-stage compost image annotation framework, leveraging an onboard automatic speech recognition model for easy-to-use unstructured image labeling and a companion mobile smartphone application for detailed instance segmentation annotation. Our annotation workflow enables the collection of a large set of coarsely labeled compost images with a subset of precise instance-level annotations, facilitating waste composition analysis without requiring users to manually maintain a food waste journal. From these annotations, our goal is to produce a dataset of high-quality, annotated images of commingled food waste, enabling the development of computer vision models for automatic food waste recognition. Using this individualized food waste data, we provide a waste analytics dashboard to users, visualizing their waste habits and encouraging sustainable consumption behavior through reflection.

Towards the development of a computer vision model for food waste recognition, we evaluate the domain transfer capabilities of YOLOE [6], an open-vocabulary object detection model. YOLOE combines the robust object localization capabilities of YOLO object detection models with CLIP, a contrastive image-language model. CLIP [7] models, with their natural language grounding and web-scale pretraining, facilitate classification over an open set of possible categorical labels from unstructured image-text labels, such as those provided by the transcribed speech annotations from our compost bin. The embedding vectors yielded by CLIP models are integrated into YOLO's object detection model for open-vocabulary labeling of previously unseen instances within an image. Our goal is to leverage the abundance of large, high-quality object detection and instance segmentation datasets for images of uneaten food to train an open-vocabulary food image recognition model. We demonstrate that such a model does not transfer to the post-consumer food waste image recognition domain, motivating the need for a large, high-quality annotated food waste image dataset.

2 Food waste impacts

The United Nations (UN) Sustainable Development Goals (SDG) outline critical environmental targets to coordinate an international response to the growing threats of global climate change and resource scarcity. UN SDG 12.3 declares the following: *"By 2030, halve per capita global food waste at the retail and consumer levels and reduce food losses along production and supply chains, including post-harvest losses"* [8]. Food waste and loss (FLW) is an immediate threat to global food security and climate, yielding significant economic losses while leading to significant direct and indirect greenhouse gas emissions. Residential households, the largest single food waste stream, comprise an estimated 60% of global food waste [1]. However, this waste source is also the most difficult to accurately measure, as a lack of standardized waste definitions and waste measurement systems results in inaccurate underestimates of total household food waste [3].

The United Nations Environment Programme (UNEP) is the custodian organization responsible for tracking food waste, defined as food disposed or spoiled at the retail, commercial, or residential level [1]. In 2024, the UNEP published their report on international progress towards SDG 12.3, finding that member nations have not made sufficient progress towards consistent, accurate food waste measurement, and therefore any analysis of progress towards food waste reduction is severely limited. Despite these deficits in measurement, the UNEP estimates global food waste volumes at 1.05 gigatonnes, with 631 million tonnes of this waste occurring in the household [1]. These estimates are drawn from six methods: direct measurement, waste composition analysis (WCA), volumetric assessment, mass balance, food item counting, and waste journaling (see Table 1 for a summary of these methods). Food waste estimates are aggregated for each nation by scaling estimates derived from national environmental agency measurements and academic studies. Only 20% of recent academic household food waste measurement studies involve direct measurement, many instead relying on indirect estimates of household food disposals [3]. The proportion and total volume of global household food waste, therefore, are poorly understood, and have been consistently underestimated on both a national and international level [1, 5].

Global FLW is responsible for 8-10% of global greenhouse-gas emissions, contributing to climate change and other adverse environmental impacts [9]. Consumer food waste which is directed to the landfill decomposes at a much faster rate than most landfilled waste, resulting in large methane emissions which are unable to be mitigated by typical methane capture systems [10]. Indirectly, food waste is responsible for greenhouse gas emissions from the energy and resources required to produce food which is ultimately wasted. 24% of freshwater used for agriculture and 30% of agricultural land are used to produce food which goes uneaten [9, 11].

This misappropriation of food resources also contributes to global hunger. An estimated 783 million people suffer from chronic hunger, with 2.4 billion experiencing moderate or severe food insecurity [8]. This has considerable downstream impacts on public health, economic advancement, and childhood development. 22.3% of children are impacted by growth stunting due to food insecurity and malnutrition. Even when considering only edible/avoidable food waste, the total volume of this food waste could provide up to 1.3 meals per person experiencing hunger each day [1].

The cumulative effects of food loss and waste on national economies is drastic. It is estimated that FLW accounts for $1 trillion of global economic losses [8]. For US consumers, these losses are estimated at $13.27 per capita per day, with $3.62 of these losses attributed to preventable, edible food waste [12]. Reducing food waste by 50% as targeted by UN SDG 12.3 could have measurable benefits for household consumers, yielding savings equivalent to 9% of a home consumer's annual food budget [13]. FLW reduction presents a $2.3 trillion annual global market opportunity, with the possibility of creating jobs and capital through public-private partnerships to reduce food waste and redirect wasted food to more energy-efficient channels [14].

To realize this economic potential and offset the insidious environmental and social harms of food waste, we must first design effective solutions for quantifying and analyzing food waste. These solutions must be accurate

Table 1 Comparison of primary food waste measurement approaches

Study type	Procedure	Merits	Weaknesses
Mass balance	Estimate waste from proxy data	Scalable, low-cost	Highly inaccurate, insufficient granularity
Questionnaires	Participants reflect and self-assess recent waste volumes	Low-cost, low participant commitment	Users may misrepresent and underreport waste
Waste composition analysis	Researchers manually collect, sort, and measure food waste	High precision, detailed food waste classification	Very high cost, no standard definitions for food categories
Food waste diary	Participants maintain a food waste log	Reduce reliance on participant memory	High participant burden, underreported waste
Direct weighing	Participants weigh waste before disposal	Precise waste volume measurement	Moderate participant burden, lacks categorical data
Waste imaging	Participants take photos of waste	Accurate waste composition data, prevents misreporting	Requires complex CV models, lacks mass estimation

while providing detailed composition analysis of wasted food in order to benefit researchers and policymakers in their efforts to design sustainable interventions.

2.1 Mitigating food waste

Attaining accurate and detailed household food waste measurements is crucial to encourage sustainable consumer behavior. Papargyropoulou et. al. defines a hierarchy of food waste disposal behaviors [15]. Food waste is best mitigated by avoiding the waste entirely, either by preventing overallocation or redistributing food resources for re-use. Composting repurposes wasted food for agricultural use. Encouraging household consumers to compost their food waste does not prevent environmental harms from wasted food, but instead allows for the food to re-enter the agricultural lifecycle as nutrients for growing new food, mitigating greenhouse gas emissions caused by the landfilling of food waste [16]. Despite this, the U.S. EPA estimates only 3.7% of U.S. residential food waste is composted [17]. Optimally, food waste measurement systems would differentiate avoidable and inevitable food waste, directing users towards sustainable habits to mitigate avoidable food waste while promoting the composting of inevitable food waste.

Household food waste drivers can be broadly classified into one of the following wasteful behaviors: household planning failures, commercial incentives to over-provision, food storage and preparation waste, unconsumed food, and disposal inefficiencies [18]. Consumer food waste should not be considered as a single behavioral failure on the part of the individual consumer. Instead, consumer food waste is a result of systemic inefficiencies in food distribution and utilization [19]. Consumers are often placed in positions in which the economically optimal solution, in terms of both financial and labor costs, is to discard food, as the utility cost of reusing that food item would be too high [20]. To mitigate this waste, users can be encouraged to adopt sustainable consumption habits. Behavioral nudges can motivate users to modify their food consumption patterns [21]; however such suggestions must be tailored to the user's waste behavior and cultural norms [22, 23], further motivating the need for individualized waste quantification rather than aggregate measures.

Directly measuring individual households' food waste both provides data to drive waste reduction interventions and serves as an intervention itself [24]. By informing consumers of the real quantity of their waste, it may motivate them to make more sustainable decisions with regard to food provisioning, preparation, and consumption. A lack of technological solutions for household consumers has been identified as a key limiting factor to household food waste reduction. In response, sustainability researchers call for "affordable technology and tools (e.g., smart kitchen tools) to improve food management" [23]. Such technologies have seen some initial adoption in institutional kitchens and been demonstrated to be effective at reducing wasteful behaviors [25].

3 Food waste measurement

Data scarcity and a lack of standardization in waste measurement are key limiting factors to reducing global food waste [26]. A growing number of food waste measurement studies aiming to address this lack of data. However, these approaches tend to rely on indirect measurements of food waste, such as mass balancing or self-reported waste measures. These methods tend to underreport actual waste quantities [4]; however 80% of recent food waste measurement studies rely on such methods [2]. Such studies disproportionately survey high-income nations, with a particular lack of data on consumer waste behaviors in developing nations. Among the few studies which do rely on direct measures of food waste, typically involve manual sorting and weighing of post-consumer waste, which is expensive, time-consuming, and potentially hazardous. Additionally, these studies are highly variable and incomparable, due to a lack of standardized definitions of food waste [3].

3.1 Waste measurement methods

Despite comprising 60% of total food waste, household food waste is the most frequently underestimated and least understood waste stream [2]. There has been an increasing interest among sustainability researchers to attain accurate and representative data of household food waste streams [3]. Existing methods for measuring household food waste can be broken down into one of six categories, as shown in Table 1.

Due to the significant cost of precise waste measurement, food waste is often inferred indirectly. Mass balance estimates of food waste are calculated using preexisting data, taking the total volumes of food inventory in an area and measuring that against average food disposal and consumption volumes to approximate a food waste rate for some region [27–29]. Questionnaire methods similarly rely on indirect waste measures, asking participants to reflect on their waste habits and report what they believe to be their total waste volumes. Such approaches do not provide sufficient granularity [1] and often do not correlate to direct waste measurements [30, 31].

Direct measurements quantify food waste at time of disposal. In household food waste studies, users are typically tasked with maintaining a written record of each discarded food item [32, 33]. These food waste diaries have been found to report much higher volumes of waste than self-assessment questionnaires [34], though they may still not accurately represent waste volumes [4], possibly due to significant annotator burden [35]. To reduce this burden, prior approaches have requested users to only record the total weight of their food waste, rather than itemizing each disposal [36]. However, reducing the granularity of food waste data inhibits the efficacy of targeted waste reduction interventions [37]. Self-reported waste measurements, including those including weight measurement, underestimate actual food waste by anywhere from 7% to 40% of real waste volume [38]. In these self-assessed measurement studies, social desirability bias and selection bias may result in the sample population misrepresenting the target data [38, 39].

To mitigate these biases, researchers have conducted waste measurement studies where researchers manually sort and quantify participants' disposals, requiring no behavioral change from the user. These waste composition analysis (WCA) studies enable researchers to attain accurate categorical data on a household's food waste [40], however, the significant labor overhead in manual waste sorting prevents this approach from being applied at scale. Additionally, cultural norms and inconsistent definitions as to what defines avoidable waste or inedible food, as well as varying descriptors used to define certain foods and varying levels of granularity in food categorization (e.g. "vegetable" vs. "green onion") lead to most WCA studies failing to be statistically representative of a target population's waste behavior [36, 41].

3.2 Technological approaches to waste measurement

Existing methods to measure household food waste fall short of delivering accurate and scalable estimates. Precise WCA approaches are too labor-intensive to be implemented broadly, while low-cost questionnaires underrepresent waste volumes. Self-assessed direct waste measurements balance cost and accuracy, but misreporting due to annotator burden yields a misrepresentation of food waste data [38]. In response, researchers have attempted to leverage technological solutions to reduce this annotator burden and improve the accuracy of household-level food waste measurement systems.

Imaging-based approaches to food measurement have garnered significant research interest, particularly for nutritional tracking [42–44] and agricultural [45, 46] purposes. Initial attempts to extend the use of camera systems to food waste measurement have shown promise, with photographic measures correlating strongly with laboratory-measured waste [4, 47, 48]. Images of household food waste can be captured by a user's smartphone [4, 49–51], enabling waste data collection at scale. Users report these systems to be less burdensome and time-consuming than maintaining a food waste diary [52]. Direct weight measurement can be used in tandem with waste photography methods to record accurate mass density and waste types [48].

Many existing household food waste imaging approaches share significant shortcomings. Many studies omit weight measurement entirely or require a two-step process where waste is weighed and imaged separately. There

have been attempts at simplifying this process by estimating waste volumes from waste images [53, 54]; however, such methods do not attain sufficient accuracy to replace direct food mass measurement. Additionally, these proposed methods rely on waste image classification to be performed by researchers [51] or users [50]. This image labeling process is an expensive and time-consuming endeavor, limiting the scale at which imaging-based waste measurement solutions can be deployed.

Dedicated food waste measurement appliances, designed for commercial food waste management, have enabled accurate waste measurement while minimizing annotation labor in busy kitchen environments. Initial approaches rely on mass measurement using smart kitchen scales [55, 56] to measure food preparation waste and customer plate waste. Iterations on these approaches include annotation tablets, with which staff can classify a record by menu item at measurement time [57]. For institutions with less rigid menus, overhead cameras have been added to smart kitchen scale systems to record waste images [58, 59]. The deployment of food waste measurement devices to institutional kitchens has led to waste reduction rates between 16% [59] and 64% [56], demonstrating the efficacy of waste measurement as a means of waste reduction [24]. This data can be used to generate data-driven food waste analytics, providing institutions with targeted recommendations to further reduce waste [60].

Similar technologies have been applied to the measurement of inorganic waste and common household refuse. Various prototypes for household-scale smart waste bins incorporate lid-mounted camera systems to capture images of waste at disposal time [61, 62]. Computer vision methods can be used to automatically sort and classify this waste, enabling the separation of organic and inorganic waste [61, 62] or recyclables from landfill waste [63, 64]. Smart waste bins have similarly been designed and deployed for large municipalities, using weight and environmental sensors to measure bin utilization and automate waste sorting [65]. Such bins can be combined into a robust Internet-of-Things (IoT) sensor network to aid municipal waste management authorities in waste generation prediction, waste retrieval route optimization, and precise evaluation of waste reduction interventions [65–68].

4 Food image recognition

Food waste image classification systems are limited by the lack of publicly available, high-quality food waste image datasets. However, such datasets do exist for images of edible food, and there has been substantial work towards developing computer vision models for food identification [69]. As our approach seeks to leverage food image recognition models for food waste recognition, we briefly review major food image datasets, listed in Table 2, and survey existing food image recognition models.

4.1 Datasets

Food image classification datasets contain text-image pairs, with each image categorized as a food item or dish. Often, these datasets are constructed by crawling image search engines [71] or food and cooking websites [73] as these platforms already host user-generated food images, searchable by food category. Food-101 [70], based

Table 2 List of common food image recognition datasets

Dataset	Year	Data type	# Images	# Classes
Food-101 [70]	2014	Classification	101,000	101
Food-201 [44]	2015	Segmentation	12,000	201
FoodX-251 [71]	2019	Classification	118,000	251
ISIA Food-500 [72]	2020	Classification	399,726	500
Recipe-1 M+ [73]	2021	Image Caption	13,765,679	N/A
FoodSeg103 [74]	2021	Segmentation	9,490	103
UECFoodPix [75]	2021	Segmentation	10,000	102

on the now defunct website `foodspotting.com`, collects the first 1000 images for the top 101 most popular food categories from this platform. FoodX-251 [71] extends Food-101's ontology, using WordNet [76] to identify similar food descriptor labels, then scraping images from Google Images which match these identifiers. 18,000 of these 118,000 images were manually validated by human annotators. A fine-tuned ResNet-101 model achieves 17% top-3 accuracy on this dataset, demonstrating both the difficulty of differentiating similar food classes and the need for more robust image classification for highly multi-class problems.

The Food-500 dataset identifies a list of 500 food categories from Wikipedia, and collects nearly 400,000 matching food images from online image search platforms [72]. The authors propose SGLANet (Stacked Global–Local Attention Network), a two-stage transformer-based model which extracts both local and global spatial features, which achieves an 89% top-5 accuracy on the Food-500 dataset.

For many food recognition and waste analysis tasks, researchers care about not only the food item present in the image, but the location of this food item in a complex image. In the context of waste analysis, this can enable automatic sorting systems [77] or provide waste composition analysis from commingled waste bins [62]. In the case of food image recognition, the ability to identify and segment multiple food items in a single image enables better calorie estimation and helps identify multiple dishes present on a single plate, as is common in Western dining settings [73]. Such approaches rely on instance segmentation, a class of computer vision models which identify objects within an image and separate these objects by drawing pixel-level masks around the items of interest. Each segmented item can then be assigned a categorical label.

Food image segmentation datasets provide food images paired with these segmentation masks, identifying each food item present in an image. Food-201, an extension of Food-101, provides 12,000 segmented food images with 201 food item classes [44]. Bounding box annotations are provided by crowdsourced annotators, and segmentation annotations are generated using GrabCut [78], an algorithm for automatic foreground object segmentation. Each image contains one or more food item labels. Similar datasets, such as UECFoodPix [75] and FoodSeg103 [74], each provide food image segmentation masks manually drawn by human annotators, rather than automated boundaries drawn by GrabCut.

4.2 Food image recognition models

Various computer vision methods have been proposed for the task of food image classification, including support vector machines using hand-crafted features [79], convolutional neural networks (CNNs) [43, 44, 80], and vision transformers [81]. Food image recognition is often applied to calorie estimation tasks for nutrition monitoring [82]. These systems generally rely on users' smartphones for data entry and imaging, providing a convenient platform for automatic caloric intake measurement.

CNNs remain the most popular approach for food image recognition tasks, with 58% of recent food image recognition studies utilizing a CNN for food recognition [83].[2] Broadly, these systems are trained on publicly available food datasets, and fine-tuned on an annotated subset of food images [44, 80, 83], with many early models being pre-trained on the ImageNet [84] dataset of images [43, 44, 80], which, incidentally, contains a set of at least 1000 food categories [85].

Food recognition models can be broadly categorized into image classification [80, 85], object detection [79, 86–88], and instance segmentation [44] models. Image classification tasks simply apply a categorical label to an image, and are only accurate in images which contain a single food entity (e.g., a dish or single ingredient). Many food image recognition tasks involve the classification of multiple food entities on a single plate, as is the case in cafeteria tray images [86, 88] or ingredient recognition in plated food images [73]. For these tasks, object detection and instance segmentation models provide granular multiclass food recognition in complex scenes.

[2] The food recognition systems survey conducted by Dalakleidi et. al, 2022 [83], was presented at a time before the widespread adoption of novel computer vision methods such as vision transformers or contrastive image embedding models, and this proportion may have changed since the study's publication in 2022.

Modern approaches to food object detection tasks often rely on YOLOv8 [89], a member of the You-Only-Look-Once (YOLO) family of single-shot object detection CNN models pre-trained on the COCO dataset [90]. Such models are often fine-tuned with only 1,000 to 20,000 annotated images, a relatively small training set given the complexity of the task, and attain a mAP@50[3] of 0.67 [86] to 0.70 [87], indicating this model's robust capability for adaptation to the downstream task of food instance localization in complex scenes. YOLOv8 has also been evaluated in other downstream food image recognition tasks, including plate waste measurement in cafeteria environments [91] and damaged food detection for food processing quality control [92]. Further applications of YOLO models have been demonstrated across a variety of solid waste image recognition tasks [93, 94].

Convolutional neural networks, though very powerful, fail to encapsulate the complex relationships between food categories. As each image is classified into a separate category, the model retains no knowledge of similarity between classes. In food descriptors, this is a key challenge, as a lack of standardized food ontology has limited the reproducibility of prior, manual food waste annotation studies [36]. Contrastive image-language models enable the multimodal fusion of language data and visual representations. Models such as CLIP train a Transformer text encoder model to represent language information, and a vision transformer or CNN for image encoding [7]. The encoders are fine-tuned to align text descriptors with matching image data, enabling the model to learn a unified embedding representation. Using these multimodal embeddings enables zero-shot classification, allowing a model to classify a sample belonging to a previously unseen category. In the context of food waste recognition, in which there exist hundreds of thousands of potential unique classes, and the same visual input may be given disparate labels by different users (e.g., "green onion" vs. "scallion"), CLIP-style models show potential for robust performance. When evaluated on Food-101, CLIP-ViT attains 95.4% classification accuracy [7].

5 Smart compost bin

To address the limitations of existing methods for household food waste measurement, we design an AI-assisted household compost bin for the measurement of residential food waste [95]. Our device implements an array of sensors to capture information about the visual, physical, and chemical qualities of each disposal. When combined with a robust compost image recognition model, this device will facilitate automatic quantification and classification of food waste items, with the goal of providing a similar level of granularity as waste composition analysis studies with minimal human input. To drive the development of such a model, we implement an easy-to-use dictation-based waste annotation framework, building a dataset of labeled food waste images to train and iteratively refine our compost image recognition model. This device will enable sustainability policymakers and environmental scientists to accurately measure household food waste streams, providing detailed insights on current waste rates and facilitating the evaluation of targeted food waste reduction efforts towards UN SDG 12.3.

5.1 Hardware design

The smart compost bin, shown in Fig. 1, is a 3D-printed device with a hinged lid and removable polycarbonate waste receptacle. Our expectation is that users place the bin on their kitchen counter to temporarily hold food waste until its later disposal in an outdoor compost bin. As the bin will sit in a user's home, near their food preparation area, we design the bin with ease of use and sanitation in mind. We consider the following four core usability principles:

[3] Mean Average Precision (mAP) is a measure of an object detection model's ability to correctly localize an object of interest in an image. In object detection model evaluation, each predicted bounding box is classified as correct if some percentage of the predicted bounding box overlaps the with the ground truth bounding box. CV practitioners refer to this overlap as intersection-over-union (IoU). mAP@50 refers to the average precision, taken as a mean over each class, with an IoU threshold of 0.50, meaning at least 50% of a predicted bounding box must overlap with the ground truth to be classified as a correct prediction.



(a) Interior (Render) (b) Bottom (Render) (c) Front (Final) (d) Rear (Final)

Fig. 1 The smart compost bin, our household food waste measurement device



(a) Baseplate (b) Body (back) (c) Main Body (d) Pi Mount

(e) Hinge (f) Camera (g) Lid Assembly (h) Lid Cutaway

Fig. 2 3D renders of the main components of the smart compost bin

Sanitation: The device must enable the user to safely dispose of their waste and clean any resulting contamination. Waste is contained in a removable and dishwasher-safe inner liner, enabling easy disposal and cleaning.
Odor Mitigation: The device must mitigate offensive odors typically emitted from food waste. The lid of the bin features a passive air filtration system, directing emissions through a carbon pellet filter to capture volatile organic compounds. A rubber gasket maintains a seal when the lid is closed.
Convenience: Users should be able to quickly interact with the device while cooking or performing household tasks. Our bin features a spring-loaded hinge and hands-free latch, allowing users to deposit refuse while their hands are occupied.
Streamlined Data Collection: First-stage data annotation should be convenient and hands-free, reducing annotator burden. The bin records the user's speech and transcribes audio into a text label using an onboard speech-to-text model.

The bin is made primarily of polylactic acid (PLA) and polyethylene terephthalate glycol (PETG) plastics. A C2-grade load cell is placed on the bottom of the bin to precisely measure the mass of each disposal (Fig. 2a and b). The main body (Fig. 2c) holds the internal disposal bin, with the Raspberry Pi 5 microcontroller and audio adapter mounted in a separate protected enclosure in the rear (Fig. 2d). The lid, which houses imaging and gas sensors, is connected to the main body using a spring-loaded hinge (Fig. 2e). An Intel RealSense camera is selected for our imaging task, providing stereoscopic image pairs which can be used for food volume estimation [96]. An infrared dot-matrix camera and a BME688 gas sensor (Fig. 2f) are also present in the lid. The lid assembly isolates the sensitive electronics from the waste chamber, preventing contamination and corrosion due to condensate. The color camera is protected from condensate and refuse by a polycarbonate cover, while the

thermal camera is shielded by an infrared-transparent plastic. These imaging sensors are positioned in the lid such that they face towards the waste receptacle when the lid is closed (Fig. 2g). Written instructions provided alongside the bin request users to periodically inspect the camera lens and, if necessary, wipe any contaminants off the polycarbonate cover with a sanitary cloth. A summary of the sensors included in the smart compost bin is provided in Table 3. An air duct allows emissions from the waste to first pass by our gas sensor, then pass through our passive carbon pellet filtration system before escaping the bin through ventilation holes. The device uses a 120 V AC power adapter to receive power from standard residential wall outlets. Each smart compost bin has a total bill-of-materials cost of $468.

When the user opens the bin by pressing the latch, the lid gently pivots open using a tensioned spring. The user then deposits their waste item and shut the lid, with the latch re-engaging to lock the lid in the closed position. The bin detects lid closure with a magnetic Hall-effect sensor, after which LEDs will illuminate the interior of the waste bin. A stereoscopic color image pair and a thermal image will be captured alongside ambient temperature and humidity, total volatile organic compounds (tVOC), and the mass of the deposited item.

We use a speech-based annotation workflow, which has been suggested to reduce annotator burden in food waste measurement studies [97]. During the data collection procedure, the bin will tell the user to describe the disposal, using the voice prompt “Please describe this item". We record the user’s speech for four seconds, during which they should describe the wasted food. An audible chime indicates to the user when this recording has ended. This speech data is processed on-board the smart compost bin’s Raspberry Pi, using an open-source C++ implementation[4] of the Whisper automatic speech recognition model [98].

5.2 Data pipeline

To facilitate the deployment of a distributed set of smart compost bins for a large-scale household food waste data collection study, we implement an integrated software system for data collection and annotation, summarized in Fig. 3. Our system consists of initial data collection and preprocessing onboard the bin’s Raspberry Pi, a central backend server for data ingress and storage, and a mobile smartphone application for data visualization and annotation (see Sect. 5.3).

Each bin is connected to a user’s home Wi-Fi network using our companion smartphone app (see Sect. 5.3). After each disposal, the compost bin takes the aggregate image and sensor data, alongside the text transcription from the user’s spoken label, and prepares a payload to be uploaded to our central database server. Data processing and upload happen asynchronously, allowing the user to initiate another food waste disposal routine immediately after the recording window for the previous scan ends. These data payloads are written to the Raspberry Pi’s storage and queued for upload to the API server. In cases where internet connectivity has been lost, the device will continue to operate, maintaining the payload queue on disk until a connection to the API server has been

Table 3 List of electronics and sensors included in the design

Component model	Purpose
Raspberry Pi 5 4GB	Connects to sensors to capture data
Intel RealSense D401	Captures stereoscopic image pair
Bosch BME688	Measures temperature, humidity, and tVOC
NAU7802 24-bit ADC	Measures signal from load cell
110° MLX90640 IR Camera	Generates thermal image and heatmap
Waveshare USB Sound Card	Plays audio prompt, records user response
US5881LUA Hall Effect Sensor	Detects when the lid is opened and closed
10 kg C2 Single-Point Load Cell	Measures mass of food items

[4] https://github.com/ggml-org/whisper.cpp.

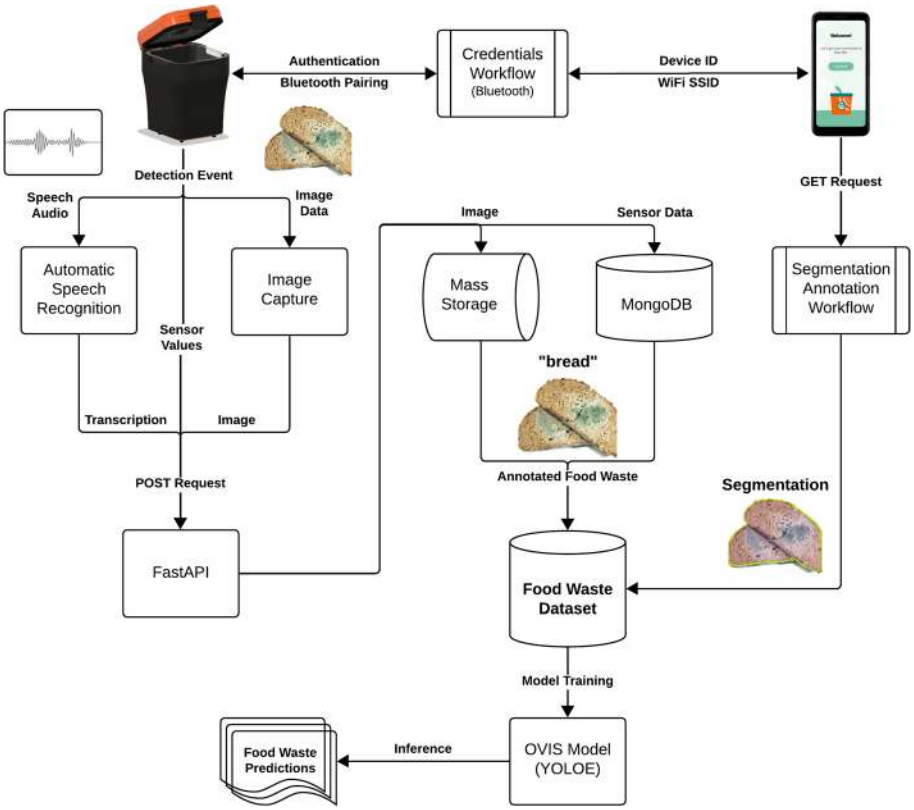
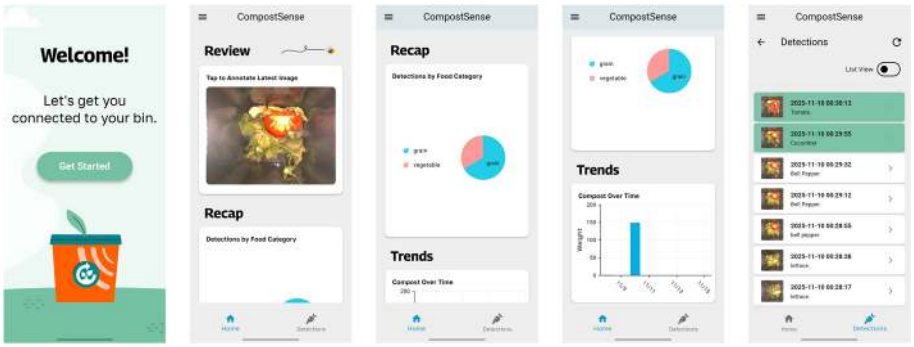


Fig. 3 System diagram detailing data collection, annotation, and storage workflow for the smart compost bin. Images are captured from the compost bin, annotated via the CompostSense mobile app, and stored in our central API server



(a) Welcome (b) Home Screen (c) Waste Group (d) Waste Mass (e) Detections

Fig. 4 User interface for the CompostSense mobile app

re-established. Once data has been successfully uploaded, the LEDs in the bin will flash green to indicate to the user that the data collection is complete.

Annotated compost records are stored on a centralized database server, with each compost bin client interfacing with our backend system via a REST API to protect our database from malformed or malicious operations. The bins send a POST request to the API to upload a compost scan, and the API will validate the format and integrity of this data. Images are written to disk, while sensor data and logging information are written to MongoDB, a non-relational (NoSQL) database. When a user opens the CompostSense mobile app, that app will request from the API the most recent images associated with that user's compost bin. The user can provide image segmentation annotations as demonstrated in Fig. 4b. If provided, the mobile client will upload these segmentations and categorical labels to the API server, where the detailed annotation is appended to the associated compost scan's database record. Image processing and analysis will be performed using a computer vision model hosted on a remote high-performance compute cluster.

5.3 CompostSense app

We design CompostSense, a companion application to the smart compost bin, as a cross-platform mobile app. The app provides a convenient interface for users to interact with their compost bin, improving our food waste measurement protocol by providing the following functionalities:

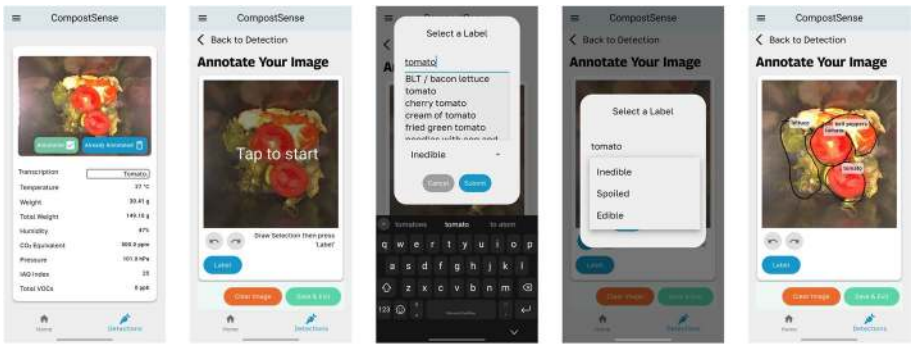
Configuration & Setup: The app allows users to set up their compost bin, connect it to their home Wi-Fi network, and reset the device in case of errors.
Detailed Image Annotations: Using the smartphone's touch input, users can provide instance-level compost image annotations, which we use to refine our food waste object localization model (see Sect. 6.1).
Waste Monitoring Analytics: CompostSense informs users of waste habits and total waste volumes over time through an analytics dashboard.

On first launch, the CompostSense app will guide the user through initial setup of their bin (see Fig. 4a). The app will initiate Bluetooth pairing with the user's compost bin, and prompt the user to provide a Wi-Fi SSID and credentials. Once the bin is connected to the user's home network, the app will exit the first-time setup and load the compost analytics dashboard.

This dashboard (see Fig. 4e) informs users about their general food waste trends. Waste items are categorized into food groups (e.g., fruits, nuts, meats) and the user is presented with a pie chart detailing the proportions of their waste by food group. Waste trends over time are given by total waste mass, as measured by the weight sensor built into the smart compost bin. This dashboard displays the most recent image taken from the bin, nudging the user towards an annotation task.

On the Detections tab (see Fig. 4d), users can select a compost scan and view details about the sensor data from this scan (see Fig. 4c). This panel will display a text transcription of the user's spoken label. We allow users to edit this label to correct mistakes from accidental misinputs or erroneous transcriptions by our automatic speech recognition model. From here, the user can tap the compost image to enter a detailed image annotation interface, from which we collect fine-grained instance segmentation labels.

We augment our coarsely labeled image dataset with fine-grained, localized annotations to iteratively refine our computer vision model's ability to separate food waste instances from commingled waste images. In the CompostSense annotation interface (see Fig. 5), users are asked to circle each food item present in the image and assign a categorical label to this object (see Fig. 4b). This interface also allows users to mark each food item as spoiled, inedible, or edible, allowing the quantification of avoidable food waste relative to the overall waste output.



(a) Waste Scan (b) Labeling (c) Category (d) State Tag (e) Annotation

Fig. 5 Compost image annotation workflow in the CompostSense mobile app

6 Open-vocabulary instance segmentation for food image recognition

Our proposed device in Sect. 5 iterates on existing household food waste measurement systems by simultaneously weighing and imaging waste, enabling users to quickly label wasted food items as part of their standard kitchen routine. This approach is an improvement on existing food waste measurement protocols, but still requires some amount of human input in the form of dictated labels. An entirely automatic system for food waste measurement would improve user continuance intention [4] and enable robust waste behavior analytics without the need for human annotation. Various imaging-based approaches to household food waste measurement have been evaluated (see Sect. 3.2), however, none provide a method for automatic waste quantification and classification from images taken in a commingled waste setting (e.g., a waste bin).

Images captured from the smart compost bin depict many possible food waste items in a complex, commingled scene. This computer vision task is challenging, as there is a need both for precise localization of food waste objects within an image and for open-vocabulary, fine-grained instance classification to capture the many possible types of food a user may dispose of. Though modern models can be trained to recognize hundreds of unique categorical labels, these approaches are still limited by the use of a fixed label set. The descriptive labels used to identify food items can vary significantly, with potentially hundreds of thousands of terms used to describe food and different terms for describing the same food item. Traditional image recognition models struggle to attain robust classification accuracy as the number of potential categories grows [99], motivating the need for an open-vocabulary food image recognition model.

We propose the use of YOLOE [6], an open-vocabulary instance segmentation model, to segment and classify food waste items using images captured from the smart compost bin. Food items placed in a compost bin will generally have significant visual overlap with other instances, resulting in many classes occupying the same spatial region in an image. Instance segmentation models predict pixel-level masks rather than bounding boxes, enabling precise localization in these visually complex scenes. YOLOE leverages the same detection pipeline as conventional YOLO object detection models (see Sect. 6.1), but adds a mask prediction branch to enable fine-grained instance segmentation. Additionally, YOLOE leverages text embeddings from CLIP, a contrastive image-language embedding model [7], to provide text-prompted open-vocabulary classification capabilities.

To date, there exists no publicly available dataset of annotated food waste images. In response, we rely on datasets of uneaten food images (see Sect. 6.2.1) to train a computer vision model for food image recognition with the goal of evaluating the transferability of food image recognition models to compost image recognition tasks. We evaluate our open-vocabulary instance segmentation model first on these datasets of uneaten food, then attempt to transfer this model to our compost instance segmentation task. This evaluation is performed using a small dataset of annotated food waste images captured from the smart compost bin during preliminary trials, which we denote FW-Test.

We train a YOLOE open-vocabulary instance segmentation model on a combination of three food image datasets: Food201 [44], UECFoodPix [75], and FoodSeg103 [74]. We compare this open-vocabulary model to `YOLO26x-seg`, a state-of-the-art closed-set instance segmentation model, to evaluate the capabilities of promptable open-vocabulary models compared to conventional image segmentation models in food recognition tasks. We assess the zero-shot performance of YOLOE on food and food waste image recognition tasks, finding that fine-tuning significantly improves model accuracy in food recognition tasks when compared to text prompting. We evaluate both zero-shot and fine-tuned YOLOE models on FW-Test, determining that training on uneaten food images alone is insufficient for an accurate compost recognition model, highlighting the need for a large-scale annotated compost image dataset.

6.1 Model architecture

YOLOE [6] is based on the popular YOLO object detection model. It utilizes a CSP-Darknet convolutional neural network (CNN) backbone for feature extraction, aggregating features from intermediate stages of this

convolutional backbone for multi-scale fusion. The feature aggregation stage, referred to as a Path Aggregation Network (PAN), leverages self-attention to improve global context awareness in downstream feature representations. Similar to `YOLO26x-seg`, YOLOE predicts anchor points for each region of interest; a separate regression head then detects bounding box coordinates while a segmentation head predicts mask coefficients to generate a pixel-wise mask. Rather than predicting a categorical value, YOLO's detection head outputs an object embedding, which is aligned with CLIP text embeddings during training to project detections into a joint region-text space.

CLIP models use jointly trained text and image encoders, learning a multimodal feature space by aligning the output embeddings from these encoders [7]. These models are often pretrained using an image captioning task, in which the goal is to maximize the cosine similarity of embedding vectors from matching text-image pairs. By grounding the image feature space in natural language context, CLIP models demonstrate strong zero-shot image classification performance, enabling these models to classify images belonging to classes not seen in the training data. The use of a separate text encoder allows for an open set of labels to be provided for an image, rather than being restricted to a fixed set of categorical labels.

YOLOE improves upon `YOLO26x-seg` by leveraging CLIP text embeddings to enable robust open-vocabulary instance classification. Rather than performing the computationally expensive task of directly fine-tuning CLIP for object-text alignment, YOLOE implements a re-parameterizable region-text alignment network (RepRTA) to adapt CLIP text embeddings to an object-text space [6], rather than the image-text space learned by CLIP's contrastive pretraining objective. This lightweight adapter enables efficient fine-tuned alignment of prompt embeddings to domain-specific visual regions while leveraging the robust pretraining of CLIP.

6.2 Results

We first train a closed-set CNN model, `YOLO26x-seg`, as a baseline for our food instance segmentation experiments (see Sect. 6.2.2). The fine-tuning of YOLO models for food image recognition tasks is a common strategy well-supported by the literature [86, 87, 91, 92] and serves as an effective baseline approach. We fine-tune four instances of `YOLO26x-seg`, one on each of our three food segmentation datasets (Food201, UECFoodPix, FoodSeg103), and one trained from all three datasets. After verifying that training on this aggregate dataset yields similar performance to dataset-specific fine-tuning, we fine-tune an instance of YOLOE on this joint dataset as described in Sect. 6.2.3. We evaluate this model's capability for domain adaptation to food recognition tasks, assessing both zero-shot prediction via text prompting as well as a conventional fine-tuning approach. Finally, we evaluate YOLOE on a dataset of annotated compost images captured from the smart compost bin to measure the capability of an open-vocabulary food object detection model to measure food waste.

6.2.1 Datasets

To train our food instance segmentation models, we use a combination of Food201, UECFoodPix-Complete, and FoodSeg103. These datasets contain labeled images of food, with a categorical label and pixel mask segmenting each food instance in the scene. After combining 18 duplicate classes across these datasets, our joint dataset contains 389 labels. In each dataset, we use the author's provided training and evaluation splits as detailed in Table 4.

In order to evaluate the transferability of our food object detection model to compost object detection tasks, we create a dataset of 467 images collected from six smart compost bins. Each bin was deployed with a beta tester from separate households for one to four months. From these 467 images, we annotate 1600 instances across 77 classes. Image labels were transcribed from user speech at time of disposal and manually reviewed for consistency. Duplicate images or those of an empty compost bin are manually filtered. Class names were normalized for capitalization ("Apple" vs. "apple") and pluralization ("grape" vs. "grapes"), but state descriptors ("apple slice" vs. "apple core") were preserved. For this evaluation subset, we rely on manual annotation, with users localizing each instance with bounding boxes. Pixel-level instance segmentations are generated from these bounding box

Table 4 Summary of food image datasets

"Unified Food Dataset" (UFD) is the union of Food201, UEC-FoodPix, and FoodSeg103. We create the Food Waste Test (FW-Test) dataset from annotated images taken from the smart compost bin

		Training		Evaluation	
Dataset	# Classes	# Samples	# Instances	# Samples	# Instances
Food object detection					
Food201 [44]	201	9,653	15,979	2,439	4,149
UECFoodPix [75]	102	9,000	13,032	1,000	1,575
FoodSeg103 [74]	103	4,983	18,318	2,135	7,697
Unified food dataset	389	23,636	47,329	5,574	13,421
Compost object detection					
FW-Test	77	–	–	467	1600

Table 5 Evaluation results of `YOLO26x-seg` instance segmentation models with dataset-specific fine-tuning

Dataset	Precision	Recall	mAP50	mAP50-95
Food201	0.478	0.312	0.335	0.247
UECFoodPix	0.798	0.575	0.627	0.534
FoodSeg103	0.481	0.389	0.375	0.308

Each model is fine-tuned on a single food segmentation dataset

Table 6 Food instance segmentation results of `YOLO26x-seg`, fine-tuned on our unified food dataset

Dataset	Precision	Recall	mAP50	mAP50-95
Food201	0.478	0.349	0.347	0.255
UECFoodPix	0.803	0.612	0.673	0.582
FoodSeg103	0.576	0.410	0.429	0.352
UFD	0.552	0.395	0.409	0.319

annotations using Segment Anything (SAM3), a zero-shot class-agnostic image segmentation model [100] and manually inspected for mask validity.

6.2.2 YOLO26 for food instance segmentation

We begin with an evaluation of a conventional CNN for food instance segmentation. We select `YOLO26x-seg` as our baseline model, fine-tuning four instances of the model on these food segmentation datasets. Each model is trained for a maximum of 80 epochs, using early stopping based on a weighted fitness criteria of $(0.1(\text{mAP50}) + 0.9(\text{mAP50-95}))$, as measured on the validation subset of each respective dataset. Each model in Table 5 is fine-tuned on a specific dataset, whereas the results shown in Table 6 demonstrate performance of a `YOLO26x-seg` instance fine-tuned on our Unified Food Dataset (UFD).

We find that fine-tuning on UFD improves model performance across all three validation sets, with improvements in mAP50 ranging from 3.58% (Food201) to 14.44% (FoodSeg103) and increases in precision up to 19.75% (FoodSeg103). These performance gains imply that, despite the increased complexity of the label space, `YOLO26x-seg` benefits from a larger training set of domain-specific examples for food image recognition tasks.

6.2.3 Open-vocabulary food instance segmentation

As YOLOE is an open-vocabulary model with text prompting, it attains generalizable zero-shot instance segmentation performance on a variety of computer vision benchmark datasets [6]. We evaluate YOLOE's zero-shot capabilities on our food and food waste benchmark tasks in Table 7. For each dataset, we provide the class labels for each category as a text prompt. YOLOE uses the CLIP embeddings of these categorical text prompts to adapt the detection model to our domain-specific ontology without requiring retraining. However, in our food domain tasks, we find that YOLOE's zero-shot functionality attains negligible performance, with an average mAP50

Table 7 Zero-shot performance of YOLOE on food segmentation datasets

Dataset	Precision	Recall	mAP50	mAP50-95
Food201	0.090	0.193	0.057	0.038
UECFoodPix	0.021	0.229	0.019	0.016
FoodSeg103	0.023	0.086	0.013	0.009
UFD	0.044	0.119	0.019	0.012

Table 8 Evaluation of YOLOE fine-tuned on Unified Food Dataset (UFD), with stock fine-tuning pipeline

Dataset	Precision	Recall	mAP50	mAP50-95
Food201	0.418	0.289	0.291	0.214
UECFoodPix	0.793	0.596	0.660	0.567
FoodSeg103	0.591	0.394	0.423	0.345
Unified food dataset	0.491	0.340	0.355	0.277

score of 0.027. However, performance on our food waste test set is substantially better, attaining a modest mAP50 score of 0.105 on our novel compost instance segmentation dataset.

To evaluate the capability of a model trained on food image data for food waste recognition tasks, we fine-tune an instance of YOLOE on UFD. This model is capable of domain adaptation post-training using text-based prompting. Categorical labels (e.g., “apple") are converted to text embeddings using CLIP, and a small lightweight adapter network is trained to align these categorical labels with the relevant visual categories represented by the segmentation masks. At inference time, the classification head is re-parameterized, representing this RepRTA network as a fixed kernel function to be convolved with the extracted vision features [6]. This re-parameterization enables efficient inference-time compute, reducing the computational overhead typically associated with open-vocabulary text-promptable image recognition models.

YOLOE’s fine-tuning pipeline performs this re-parameterization during the fine-tuning stage, freezing the text embeddings after domain adaptation. This poses a challenge for our application, in which our goal is to perform domain-specific fine-tuning on images of uneaten food, then transfer the model to food waste tasks through text prompting. YOLOE’s default fine-tuning pipeline forces the model to be “closed-set" after this initial fine-tuning on food images, prohibiting a second domain transfer.

We modify the default YOLOE training pipeline, maintaining the same pre-training configuration as specified in the Ultralytics implementation of YOLOE,[5] but not fusing the adapted prompt embeddings into the detection head. Similarly to the reference fine-tuning implementation, we freeze all weights in YOLOE with the exception of the final three convolutional blocks of the classification head. Additionally, we modify the learning rate from the default 2×10^{-3} to 1×10^{-3} to improve training stability.

We compare the performance of a YOLOE instance trained with our custom fine-tuning pipeline (Table 9) to the stock YOLOE fine-tuning pipeline (Table 8). We find that the reference implementation’s approach of keeping the text prompt embedding fixed during fine-tuning does improve in-domain performance when fine-tuning on food image data. An average decrease in mAP50 scores of 5.86% is observed across our three food benchmark datasets when text prompt embeddings are modifiable during fine-tuning.

Enabling the post-fine-tuning modification of these text prompt embeddings allows the transfer of a fine-tuned YOLOE instance to a new domain. We evaluate the performance of a YOLOE model fine-tuned on our Unified Food Dataset for food waste recognition tasks. In Table 9, we find that our finetuned YOLOE instance performs substantially worse on food waste instance segmentation tasks with the FW-Test dataset than the zero-shot results demonstrated in Table 7, with an 80% reduction in mAP50 scores after fine-tuning.

[5] https://github.com/ultralytics/ultralytics/blob/main/ultralytics/models/yolo/yoloe/train_seg.py.

Table 9 Evaluation of YOLOE fine-tuned on Unified Food Dataset using a modified fine-tuning pipeline

Dataset	Precision	Recall	mAP50	mAP50-95
Food201	0.349	0.329	0.283	0.206
UECFoodPix	0.766	0.561	0.623	0.530
FoodSeg103	0.494	0.395	0.384	0.308
Unified food dataset	0.446	0.360	0.346	0.267
FW-Test	0.010	0.205	0.021	0.016

Prompt embeddings remain modifiable during fine-tuning. Categorical labels are provided as text prompts at evaluation time, enabling the transfer of YOLOE to our food waste instance segmentation task

7 Discussion

We present three key contributions towards the development of an automatic household food waste measurement system. First, we design and manufacture a food waste measurement device with the needs of residential consumers in mind. This device facilitates a standardized food waste measurement study protocol, unifying the waste classification and weighing steps and reducing annotator burden. Our approach improves upon historically expensive and time-consuming waste composition analysis studies, seamlessly imaging each disposal and using state-of-the-art automatic speech recognition technology to allow for the dictation of waste annotations.

To support our food waste measurement device, we provide an integrated software solution for household food waste data collection and analysis. We develop a central database server and API backend to connect our fleet of IoT smart compost bins for data collation and processing. Our companion smartphone application, CompostSense, provides a platform for users to annotate food waste images with instance-level detail using the touch input of their mobile device. This app presents users with detailed information about their waste habits, including food waste volume over time and waste distribution by food group. A personalized waste behavior analytics dashboard enables users to monitor their food waste output, which may lead to personal reflections on one's waste habits and encourage the adoption of more sustainable consumption behaviors [24].

We evaluate the viability of YOLOE, a text-promptable open-vocabulary instance segmentation model, for food instance segmentation tasks towards the development of a compost image recognition system. We find that the robust pretraining of the YOLOE feature extraction backbone, combined with the generalizable natural language encoding capabilities of CLIP, enables YOLOE to attain limited zero-shot capability on our food waste recognition task. However, the subpar zero-shot performance demonstrated by YOLOE on food instance segmentation tasks exhibits the limitations of domain adaptation through text prompting alone for fine-grained food image recognition. We hypothesize this poor zero-shot transferability to be a result of a lack of annotated datasets of food waste images available online. Large vision-language models such as CLIP are often pre-trained at web-scale, with datasets aggregated using web-based image search engines [7]. However, the lack of food waste image datasets implies a limited set of compost images with instance-level labels available in these web-scale datasets. This offers one possible explanation for the poor generalizability of existing VLMs to fine-grained food waste image recognition tasks, though further analysis of a larger set of large VLMs would be beneficial to provide rigorous insight into the zero-shot food image recognition capabilities of these models.

We modify the fine-tuning pipeline of YOLOE to enable fine-tuned instances of YOLOE to have text prompts be modified at inference time. This novel approach to domain adaptation with YOLOE enables domain-specific fine-tuning on instance segmentation data, and later allows this fine-tuned model to be transferred to a new classification ontology by encoding categorical labels as text embeddings, then fusing these text embeddings into the classification head as is traditionally done during the initial fine-tuning stage with YOLOE. We find that, though waiting to fuse these text prompt embeddings until after domain-specific fine-tuning reduces in-domain task performance by an average mAP50 score of 5.86%, it enables these models to be transferred to a secondary domain after fine-tuning.

We find that the fine-tuning of a YOLOE instance on food instance segmentation data provides an open-vocabulary model which can attain in-domain performance on some food image recognition tasks comparable to a closed-set `YOLO26x-seg` model. We see a reduction in mAP50 scores of 1.93% for UECFoodPix and 1.40% for FoodSeg103, respectively. However, on Food201, which has a larger label space, we see a more significant 16.14% reduction in mAP50 scores when using YOLOE compared to `YOLO26x-seg`. These results suggest an accuracy trade-off when using an open-vocabulary instance segmentation model, even with domain-specific fine-tuning, when compared to a conventional closed-set model. We find that YOLOE's compost instance segmentation performance collapses after fine-tuning on UFD. This suggests fine-tuning on food image data alone is insufficient to develop a robust open-vocabulary compost image recognition model.

Our experiments highlight the domain gap between food and food waste image data. Post-consumer food is often highly visually dissimilar from food prepared for human consumption. Decay, state transformation (e.g. cutting, dicing), or partial consumption can all significantly visually modify a given food item. This large difference in the visual distribution of food images and food waste images which otherwise share the same categorical label may drive poor domain transfer. Our results support the poor generalizability of instance segmentation models fine-tuned for food image recognition tasks to the food waste domain in the case of the YOLO class of models. These findings motivate the need for a high-quality annotated food waste image dataset to enable fine-tuning of a model for automatic food waste segmentation and classification. The strong performance of YOLOE on food instance segmentation tasks highlights the adaptability of this model, suggesting the capability of this model architecture to be applied to future fine-grained instance segmentation tasks, such as food waste image recognition, given sufficient fine-tuning data.

7.1 Limitations

Our ability to train and evaluate food waste instance segmentation models is constrained by the size of our annotated food waste image dataset. This food waste image dataset represents data collection from six smart compost bins deployed for beta testing with volunteer users. These beta deployments were limited in scope, focusing primarily on device functionality rather than the collection of a large set of food waste images. Additionally, the sample population of beta testers is not necessarily a representative sample of household consumers. This lack of sufficient data prevents us from fine-tuning a model on compost image data, instead requiring the use of uneaten food image datasets. This prohibits a direct evaluation of the transferability of food image recognition models to compost image recognition tasks, as we lack a fine-tuned compost image recognition model to compare against.

A standardized definition for the specificity and granularity of categorical labels does not exist across food image recognition datasets. Some datasets, such as Food201 [44] and UECFoodPix [75], provide dish-level annotations, labeling each food item based on completed dishes (e.g., "beef curry" [75]), rather than itemizing each ingredient. Conversely, FoodSeg103 [74] provides ingredient-level annotations, labeling each ingredient on a plated dish as a separate instance. We hypothesize that this varying specificity introduces noise into our training procedure, in which we train with all three of the aforementioned datasets. Food waste image recognition requires a model which can perform both ingredient-level and dish-level classification. Users may dispose of an entire plated dish, such as leftovers, and label the entire dish as one disposal. The same user may then prepare a cooked meal, disposing of individual ingredients during the cooking process, introducing varying levels of granularity into our label space.

There is a considerable difference in the visual appearances of uneaten food and food waste items. Visual changes can occur due to decay, or a transformation in state as a byproduct of cooking or consumption (e.g., an apple core appearing differently from an apple). We hypothesize that these visual disparities may be a contributing factor to the failure of our food instance segmentation model to transfer to food waste recognition tasks.

Our fine-tuned food instance segmentation model is further limited by the lack of publicly available food instance segmentation datasets. Though more prevalent than food waste instance segmentation data, our unified food segmentation dataset only features 23,636 images and 389 classes. Though sufficient for in-domain

fine-tuning, this relative paucity of data prohibits the development of a model designed for broad domain coverage for a fine-grained food instance classification. As YOLOE requires instance segmentation datasets for fine-tuning, it precludes the use of the robust set of food image classification and food object detection datasets present in the literature [71, 73].

One significant challenge with commingled food waste image collection is instance-label correspondence. Our first-stage annotation workflow captures a single text label alongside an image of the entire contents of the compost bin. This label will only describe a single instance within the image, without providing the appropriate grounding to isolate that instance from other food waste items contained in the image. Our goal with these labels is to fine-tune a vision-language model for generalized food waste image recognition tasks, relying on separate instance segmentation predictors for instance localization. The instance-level annotations provided via the CompostSense app provide localized per-object labels, facilitating the training of instance segmentation models. However, attributing the verbal annotations collected at measurement time to the most recently disposed object in the image is a future task of interest for refining our data collection protocol.

7.2 Future work

To develop a robust computer vision model for automatic food waste measurement, we must first collect a sufficiently large dataset of annotated food waste images. To date, we have manufactured 55 smart compost bins, with the goal of distributing these units to volunteer households across the state of Oregon. Through the data annotation protocol detailed in Sect. 5, we will collect a large set of coarse-grained dictated waste labels as well as a small subset of instance-level annotations provided by these volunteers. Through a combination of coarse-grained image-text pairs given by user dictation and instance segmentation annotations provided through the CompostSense app, we will train future models on this dataset to provide a task-specific open-vocabulary food waste image recognition model without the need to rely on datasets of uneaten food images.

Though YOLOE's fine-tuning pipeline requires images to be annotated with segmentation masks, components of the YOLOE model could be iteratively refined with less granular food image data to improve domain transferability. The CSP-Darknet feature extraction backbone of YOLOE could be fine-tuned on food image classification datasets, such as FoodX-251 [71] or ISIA Food-500 [72] (see Table 2). YOLOE uses MobileCLIP, a lightweight image-text embedding model based on CLIP and trained on DataCompDR-1B [101], a web-scale multimodal image-text dataset. However, this text prompt embedding model could be further refined by pre-training on domain-specific image-text datasets, such as Recipe1M+ [73]. We hypothesize that fine-tuning these underlying components of YOLOE may improve the model's adaptability to food domain tasks, as these food datasets with less granular localized annotations feature significantly more images and categorical labels than are available in Food201, UECFoodPix, or FoodSeg103.

There is a growing body of research in open-vocabulary instance segmentation models. Some recent work has leveraged foundation models such as Segment Anything (SAM) in two-stage detection frameworks. In this approach, SAM provides class-agnostic segmentation proposals [100], which are used to train a modified CLIP model for contrastive segmentation-text matching [102, 103]. Single-stage end-to-end approaches utilize side adapter networks (SAN) to predict mask proposals [104] by training a lightweight vision transformer using intermediate features from CLIP to adapt CLIP for dense prediction tasks. Such approaches have been applied to food instance segmentation, where the CLIP model is augmented with a domain-specific food image feature extraction model to improve domain adaptation and fine-grained food instance segmentation while leveraging the open-vocabulary segmentation capabilities of SAN [105]. YOLOE is designed as a lightweight CNN intended for deployment on edge devices; however this design choice limits the scope and capability of the model. We intend to explore alternative open-vocabulary instance segmentation models for the task of food waste instance segmentation.

Foundation vision-language models (VLMs) are becoming increasingly common for image understanding tasks. Though many current VLMs struggle with fine-grained image classification [106], they still demonstrate

promising open-vocabulary image recognition capabilities thanks to rich natural text grounding and web-scale pretraining [107, 108]. Such models have become capable of instance localization tasks, learning to generate bounding box coordinates or tokenized segmentation masks through task-specific fine-tuning [109]. We will assess the capabilities of such models for food waste segmentation tasks, leveraging our forthcoming compost image dataset to design an instruction tuning task for the adaptation of VLMs for food waste measurement.

The multimodal embeddings learned by a CLIP model fine-tuned for food image recognition tasks can encode documents beyond just food images and descriptors, such as recipes. Recipe1M+ provides a robust set of recipes, itemized by ingredients [73]. These recipes could be retrieved by querying a vector database for documents mentioning certain ingredients. The individualized food waste measurements provided by our smart compost bin could enable targeted food waste interventions by providing recipe recommendations based on a user's frequently wasted ingredients. Targeted recommendations could be integrated into the waste analytics dashboard in our smartphone app, offering a convenient interface for food waste reduction through behavioral nudging.

Our long-term goal with the development of the smart compost bin is to provide a food waste measurement device which can quantify individual household food waste while minimizing user burden. The development of an open-vocabulary food waste image recognition model will enable wide-scale household food waste measurement and provide personalized waste analytics without the need for labor-intensive manual annotation. To achieve this, we pursue the development of a robust open-vocabulary compost instance segmentation model and the production of a high-quality annotated compost image dataset. The smart compost bin is designed with this automatic measurement routine in mind, and in the future, we aim to simply provide a software update to convert the smart compost bin from a convenient waste image annotation framework to an automatic food waste measurement appliance.

8 Conclusions

We present a 3D-printed hardware device for the measurement of household food waste, addressing a critical gap in existing food waste and loss (FLW) measurement literature. Our food waste measurement protocol enables the collection of detailed compost image annotations, contributing towards the development of a computer vision model for automatic food waste classification. Our data collection system provides a standardizable, robust household food waste measurement framework which can be deployed at scale to critically evaluate household food waste streams, facilitating the evaluation of UN member nations' progress towards SDG 12.3.

We evaluate the feasibility of a open-vocabulary instance segmentation model for fine-grained open-vocabulary food and food waste image recognition tasks. We find that YOLOE attains modest performance on our food waste instance segmentation task, and demonstrates further potential domain adaptation capabilities when fine-tuned on food domain specific data. However, the transfer of a model fine-tuned for food image recognition to food waste instance segmentation tasks results in a substantial performance loss, indicating the limitations of currently available food image datasets. The inability of food object recognition models to transfer to our food waste image recognition task highlights the need for a large dataset of annotated compost images to train a food waste object detection model. Using the data collection protocol enabled by our smart compost bin, we aim to provide such a dataset in our ongoing food waste image annotation study.

Appendix A Training configuration

We outline the exact fine-tuning configuration used for YOLOE and `yolo26x-seg` for the experiments presented in Sect. 6.2.
See Table 10.

Table 10 Hyperparameters for the training of `yolo26x-seg` and YOLOE for food instance segmentation

	YOLO26x-seg	YOLOE	
Experiment	Closed-set training	Fixed vocabulary	Prompt-enabled
Data and model			
Model checkpoint	`yolo26x-seg.pt`	`yoloe-26x-seg.pt`	`yoloe-26x-seg.pt`
Trainer	Ultralytics segmentation trainer	`YOLOEPESegTrainer`	`YOLOESegTrainerFromScratch`
Input resolution	640×640	640×640	640×640
Training epochs	80	80	80
Batch size	16	16	16
Gradient accumulation	4	4	4
Effective update batch size	64	64	64
Optimization and learning-rate schedule			
Optimizer	MuSGD[a]	MuSGD[a]	AdamW
Base initial learning rate	1.0×10^{-2}	1.0×10^{-2}	1.0×10^{-3}
Boosted MuSGD-group learning rate	3.0×10^{-2}	3.0×10^{-2}	–
Final base-group learning rate	2.2375×10^{-4}	2.2375×10^{-4}	2.2375×10^{-5}
Final boosted-group learning rate	6.7125×10^{-4}	6.7125×10^{-4}	–
Momentum / AdamW coefficients	$0.8 \rightarrow 0.937$	$0.8 \rightarrow 0.937$	$\beta_1 = 0.9,\ \beta_2 = 0.999$
Weight decay	5.0×10^{-4}	5.0×10^{-4}	2.5×10^{-2}
Box-loss weight	7.5	7.5	7.5
Classification-loss weight	0.5	0.5	0.5
Distribution-focal-loss weight	1.5	1.5	1.5
Training augmentations			
Translation	±10%	±10%	±10%
Scale variation	±50%	±50%	±50%
Horizontal-flip probability	0.5	0.5	0.5
Vertical-flip probability	0	0	0
HSV augmentation (h, s, v)	(0.015, 0.7, 0.4)	(0.015, 0.7, 0.4)	(0.015, 0.7, 0.4)
Evaluation configuration			
Evaluation resolution	640×640	640×640	640×640
Validation batch size	32	32	32
Configured NMS IoU threshold	0.7[d]	0.7[d]	0.7[d]
AP matching thresholds	0.50 : 0.05 : 0.95	0.50 : 0.05 : 0.95	0.50 : 0.05 : 0.95

[a]The optimizer was selected by Ultralytics' `optimizer=auto`; the listed optimizer and learning rate are the effective values reported in the Ultralytics execution logs

Acknowledgements The authors appreciate the efforts of Oregon State University undergraduate engineering students Daniel Eastman, Jake Enos, Vaylen Melendez, Kevin Ng, and William Richards, who contributed greatly in the development and fabrication of the smart compost bin; Daniel Lau, Kimberly Markley, Blake Pugh, and Micah Stalberg, who contributed to the development of the CompostSense app; and Vi Vuong, who assisted in the annotation of our compost image dataset.

Author contributions All authors contributed equally.

Funding This research was funded by the Foundation for Food and Agriculture Research and the Kroger Zero Hunger Zero Waste Foundation (FFAR Grant ID: 22-000420).

Data availability The dataset presented in this article is not readily available, as the dataset is pending a forthcoming human subject study and has not been finalized. Requests to access the dataset should be directed to the corresponding author at beerya@oregonstate.edu.

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

References

1. Forbes H, Peacock E, Abbot N, Jones M. Think eat save: tracking progress to halve global food waste. 2024. (Food Waste Index Report United Nations Environment Programme).
2. Xue L, Liu G, Parfitt J, Liu X, Van Herpen E, Stenmarck A, et al. Missing food, missing data? A critical review of global food losses and food waste data. Environ Sci Technol. 2017;51(12):6618–33. https://doi.org/10.1021/acs.est.7b00401.
3. Dou Z, Toth JD. Global primary data on consumer food waste: Rate and characteristics—a review. Resour Conserv Recycl. 2021;168:105332. https://doi.org/10.1016/j.resconrec.2020.105332.
4. Herpen E, Lans IA, Holthuysen N, Nijenhuis-de Vries M, Quested TE. Comparing wasted apples and oranges: an assessment of methods to measure household food waste. Waste Manage. 2019;88:71–84. https://doi.org/10.1016/j.wasman.2019.03.013.
5. Amicarelli V, Bux C. Food waste measurement toward a fair, healthy and environmental-friendly food system: a critical review. British Food J. 2021;123(8):2907–35. https://doi.org/10.1108/BFJ-07-2020-0658.
6. Wang A, Liu L, Chen H, Lin Z, Han J, Ding G. YOLOE: real-time seeing anything. In: Proceedings of the IEEE/cvf international conference on computer vision (ICCV). 2025. pp. 24591–602.
7. Radford A, Kim JW, Hallacy C, Ramesh A, Goh G, Agarwal S, Sastry G, Askell A, Mishkin P, Clark J, Krueger G, Sutskever I. Learning transferable visual models from natural language supervision. In: Meila M, Zhang T, editors. Proceedings of the 38th international conference on machine learning. Proceedings of machine learning research. 2021. vol. 139, pp. 8748–8763. https://proceedings.mlr.press/v139/radford21a
8. United Nations General Assembly, United Nations Economic and Social Council: Progress towards the Sustainable Development Goals. 2024. https://unstats.un.org/sdgs/files/report/2024/SG-SDG-Progress-Report-2024-advanced-unedited-version.pdf
9. Food and Agriculture Organization of the United Nations: food wastage footprint: impacts on natural resources. 2013. pp. 1–63.
10. Cusworth DH, Duren RM, Ayasse AK, Jiorle R, Howell K, Aubrey A, et al. Quantifying methane emissions from united states landfills. Science. 2024;383(6690):1499–504. https://doi.org/10.1126/science.adi7735.
11. Kummu MS, Moel H, Porkka M, Siebert S, Varis O, Ward PJ. Lost food, wasted resources: global food supply chain losses and their impacts on freshwater, cropland, and fertiliser use. Sci Total Environ. 2012;438:477–89. https://doi.org/10.1016/j.scitotenv.2012.08.092.
12. Conrad Z. Daily cost of consumer food wasted, inedible, and consumed in the United States, 2001–2016. Nutr J. 2020;19(1):35. https://doi.org/10.1186/s12937-020-00552-w.
13. Rutten M, Nowicki P, Bogaardt MJ, Aramyan L. Reducing Food Waste by Households and in Retail in the EU. https://edepot.wur.nl/290135
14. Adams J, Kanai S, Al-Mubarak NA, Sjoblom TS, Stuchtey M, Long J, Buddemeier P, Herweijer C, Scott L, Dowling D, Grant J, Stephenson J, Plant L, Bleischwitz R, Clarke G, Kabat P, Elkington J, Head P, Trust E, Schmidt-Traub G, Birt M, Wrigge G. Valuing the SDG prize in food and agriculture. Technical report. 2016. https://s3.amazonaws.com/aws-bsdc/Valuing-SDG-Food-Ag-Prize-Paper.pdf
15. Papargyropoulou E, Lozano R, Steinberger JK, Wright N, Ujang Z. The food waste hierarchy as a framework for the management of food surplus and food waste. J Clean Prod. 2014;76:106–15. https://doi.org/10.1016/j.jclepro.2014.04.020.
16. Kenny S, Stephenson J, Stern A, Beecher J, Morelli B, Henderson A, Chiang E, Beck A, Cashman S, Wexler E, McGaughy K, Martell A. from field to bin: the environmental impacts of U.S. food waste management pathways. 2023. https://www.epa.gov/system/files/documents/2023-10/part2_wf-pathways_report_formatted_no-appendices_508-compliant.pdf

17. United States Environmental Protection Agency: 2019 wasted food report. Technical Report EPA 530-R-23-005, U.S. Environmental Protection Agency. 2023. https://www.epa.gov/system/files/documents/2024-04/2019-wasted-food-report_508_opt_ec_4.23correction.pdf
18. Principato L, Mattia G, Di Leo A, Pratesi CA. The household wasteful behaviour framework: a systematic review of consumer food waste. Ind Mark Manage. 2021;93:641–9. https://doi.org/10.1016/j.indmarman.2020.07.010.
19. National Academies of Sciences, Engineering, and Medicine: A National Strategy to Reduce Food Waste at the Consumer Level. Washington, D.C.: National Academies Press; 2020. https://doi.org/10.17226/25876
20. Lusk JL, Ellison B. Economics of household food waste. Can J Agric Econ. 2020;68(4):379–86. https://doi.org/10.1111/cjag.12256.
21. Reisch LA, Sunstein CR, Andor MA, Doebbe FC, Meier J, Haddaway NR. Mitigating climate change via food consumption and food waste: a systematic map of behavioral interventions. J Clean Prod. 2021;279:123717. https://doi.org/10.1016/j.jclepro.2020.123717.
22. Boulet M, Stott A, Kneebone S. Which behaviours matter? Prioritising food waste reduction behaviours for targeted policy and program approaches. J Environ Manage. 2023;345:118668. https://doi.org/10.1016/j.jenvman.2023.118668.
23. Vittuari M, Herrero LG, Masotti M, Iori E, Caldeira C, Qian Z, et al. How to reduce consumer food waste at household level: A literature review on drivers and levers for behavioural change. Sustain Prod Consum. 2023;38:104–14. https://doi.org/10.1016/j.spc.2023.03.023.
24. Ramos GJ, Borges JAR, Faria Domingues CH, Herpen E. Reducing food waste by simply measuring it: insights from interventions to reduce household food waste. British Food J. 2023;126(2):812–33. https://doi.org/10.1108/BFJ-02-2023-0092.
25. Yu Y, Yi S, Nan X, Lo LY-H, Shigyo K, Xie L, Wicaksana J, Cheng K-T, Qu H. Foodwise: food waste reduction and behavior change on campus with data visualization and gamification. In: Proceedings of the 6th ACM SIGCAS/SIGCHI conference on computing and sustainable societies. COMPASS '23. New York, NY, USA: Association for Computing Machinery; 2023. pp. 76–83. https://doi.org/10.1145/3588001.3609364
26. Food and Agriculture Organization of the United Nations: Tracking progress on food and agriculture-related SDG indicators 2021: a report on the indicators under FAO custodianship. 2021. https://doi.org/10.4060/cb6872en
27. Buzby JC, Farah-Wells H, Hyman J. The estimated amount, value, and calories of postharvest food losses at the retail and consumer levels in the United States. USDA-ERS Econ Inf Bull. 2014. https://doi.org/10.2139/ssrn.2501659.
28. Gustafsson J, Cederberg C, Sonesson U, Emanuelsson A. The methodology of the FAO study: global food losses and food waste—extent, causes and prevention. 2013. https://api.semanticscholar.org/CorpusID:131784963
29. Den Bos VM, Vreede L, Achterbosch T, Rutten MM. Consumers discard a lot more food than widely believed: estimates of global food waste using an energy gap approach and affluence elasticity of food waste. PLoS ONE. 2020;15(2):e0228369. https://doi.org/10.1371/journal.pone.0228369.
30. Shu Y, Roe BE, Bender K. Adapting, refining and expanding a validated questionnaire to measure food waste in US households. MethodsX. 2021;8:101377. https://doi.org/10.1016/j.mex.2021.101377.
31. Elimelech E, Ert E, Ayalon O. Bridging the gap between self-assessments and measured household food waste: a hybrid valuation approach. Waste Manage. 2019;95:259–70. https://doi.org/10.1016/j.wasman.2019.06.015.
32. Richter B, Bokelmann W. Explorative study about the analysis of storing, purchasing and wasting food by using household diaries. Resour Conserv Recycl. 2017;125:181–7.
33. Boer J, Skiba A, Boer E, Obersteiner G, Dyjakon A. Prevention and under-reporting effects of food waste diaries. Sustainability. 2024;16:2411009. https://doi.org/10.3390/su162411009.
34. Giordano C, Alboni F, Falasconi L. Quantities, determinants, and awareness of households' food waste in Italy: a comparison between diary and questionnaires quantities. Sustainability. 2019;11(1212):3381. https://doi.org/10.3390/su11123381.
35. Sheble L, Wildemuth B. Applications of social research methods to questions in information and library science. Westport, CT: Bloomsbury Libraries Unlimited; 2009. pp. 211–21.
36. Withanage SV, Dias GM, Habib K. Review of household food waste quantification methods: focus on composition analysis. J Clean Prod. 2021;279:123722. https://doi.org/10.1016/j.jclepro.2020.123722.
37. Ananda J, Karunasena GG, Pearson D. Identifying interventions to reduce household food waste based on food categories. Food Policy. 2022;111:102324. https://doi.org/10.1016/j.foodpol.2022.102324.
38. Quested TE, Palmer G, Moreno LC, McDermott C, Schumacher K. Comparing diaries and waste compositional analysis for measuring food waste in the home. J Clean Prod. 2020;262:121263. https://doi.org/10.1016/j.jclepro.2020.121263.
39. Scollon CN, Kim-Prieto C, Diener E. Experience sampling: promises and pitfalls, strengths and weaknesses. J Happiness Stud. 2003;4(1):5–34. https://doi.org/10.1023/A:1023605205115.
40. Dahlén L, Lagerkvist A. Methods for household waste composition studies. Waste Manage. 2008;28(7):1100–12. https://doi.org/10.1016/j.wasman.2007.08.014.
41. Lebersorger S, Schneider F. Discussion on the methodology for determining food waste in household waste composition studies. Waste Manag (New York NY). 2011;31(9–10):1924–33. https://doi.org/10.1016/j.wasman.2011.05.023.

42. Ming Z-Y, Chen J, Cao Y, Forde C, Ngo C-W, Chua TS. Food photo recognition for dietary tracking: System and experiment. In: MultiMedia Modeling. Cham, Switzerland: Springer; 2018. pp. 129–141. https://doi.org/10.1007/978-3-319-73600-6_12
43. Ege T, Yanai K. Estimating food calories for multiple-dish food photos. In: 2017 4th IAPR asian conference on pattern recognition (ACPR). 2017. pp. 646–51.https://doi.org/10.1109/ACPR.2017.145
44. Myers A, Johnston N, Rathod V, Korattikara A, Gorban A, Silberman N, Guadarrama S, Papandreou G, Huang J, Murphy K. Im2calories: towards an automated mobile vision food diary. In: IEEE international conference on computer vision (ICCV). 2015. pp. 1233–241. https://doi.org/10.1109/ICCV.2015.146
45. Patrício DI, Rieder R. Computer vision and artificial intelligence in precision agriculture for grain crops: a systematic review. Comput Electron Agric. 2018;153:69–81. https://doi.org/10.1016/j.compag.2018.08.001.
46. Elbasi E, Mostafa N, AlArnaout Z, Zreikat AI, Cina E, Varghese G, et al. Artificial intelligence technology in the agricultural sector: a systematic literature review. IEEE Access. 2023;11:171–202. https://doi.org/10.1109/ACCESS.2022.3232485.
47. Martin CK, Han H, Coulon SM, Allen HR, Champagne CM, Anton SD. A novel method to remotely measure food intake of free-living people in real-time. Br J Nutr. 2009;101(3):446–56. https://doi.org/10.1017/S0007114508027438.
48. Cropley M, Sprajcer M, Dawson D. Wastogram: validation of a new tool to measure household food waste. J Environ Psychol. 2022;84:101896. https://doi.org/10.1016/j.jenvp.2022.101896.
49. Roe BE, Apolzan JW, Qi D, Allen HR, Martin CK. Plate waste of adults in the United States measured in free-living conditions. PLoS ONE. 2018;13(2):e0191813. https://doi.org/10.1371/journal.pone.0191813.
50. Danible K, Panizza C, Boushey C, Kerr D, Zhu F, Banna JC. A novel to method to measure food waste: the mobile food record. J Ext. 2021;59(3):1. https://doi.org/10.34068/joe.59.03.01.
51. Ganglbauer E, Fitzpatrick G, Güldenpfennig F. Why and what did we throw out? Probing on reflection through the food waste diary. In: Proceedings of the 33rd annual ACM conference on human factors in computing systems. CHI '15. New York, NY, USA: Association for Computing Machinery; 2015. pp. 1105–1114. https://doi.org/10.1145/2702123.2702284
52. Roe BE, Qi D, Beyl RA, Neubig KE, Martin CK, Apolzan JW. The validity, time burden, and user satisfaction of the FoodImage™ smartphone app for food waste measurement versus diaries: a randomized crossover trial. Resour Conserv Recycl. 2020;160:104858. https://doi.org/10.1016/j.resconrec.2020.104858.
53. Lubura J, Pezo L. Food recognition and food waste estimation using convolutional neural network. Electronics. 2022;11(22):3746. https://doi.org/10.3390/electronics11223746.
54. Hassannejad H, Matrella G, Ciampolini P, Munari ID, Mordonini M, Cagnoni S. A new approach to image-based estimation of food volume. Algorithms. 2017;10(2):66. https://doi.org/10.3390/a10020066.
55. Oliveira Pontes T, Silva César A, Conejero MA, Rodrigues Deliberador L, Otávio Batalha M. Food waste measurement in a chain of industrial restaurants in Brazil. J Clean Prod. 2022;369:133351. https://doi.org/10.1016/j.jclepro.2022.133351.
56. Silvennoinen K, Nisonen S, Pietilainen O. Food waste case study and monitoring developing in finnish food services. Waste Manage. 2019;97:97–104. https://doi.org/10.1016/j.wasman.2019.07.028.
57. Goossens Y, Leverenz D, Kuntscher M. Waste-tracking tools: a business case for more sustainable and resource efficient food services. Resour Conserv Recycl Adv. 2022;15:200112. https://doi.org/10.1016/j.rcradv.2022.200112.
58. Mekhsian H, Duffy K, Zornes M. System and method for monitoring food waste. US 11854419 B2. 2006. https://patentimages.storage.googleapis.com/a7/9c/9a/512dc9b598081d/US7415375.pdf
59. Rock S. Food waste reduction in military kitchens, a tracking technology demonstration at fort Jackson. 2019. https://cfpub.epa.gov/si/si_public_file_download.cfm?_download_id=539177&Lab=NRMRL
60. Nagasharmila P, Aadhini GK, Mohanasubi V, Pramika T, Babitha Lincy R. A holistic approach to food waste management: leveraging full-stack technologies for efficient resource utilization. In: Third international conference on smart technologies, communication and robotics (STCR). 2023. vol. 1, pp. 1–5. https://doi.org/10.1109/STCR59085.2023.10397029
61. Taurer F, Wolling F, Moore J, Michahelles F. Smart trash can: easy collection of photos of organic waste in the home. In: Proceedings of the 22nd international conference on mobile and ubiquitous multimedia. New York, NY, USA: Association for Computing Machinery; 2023. pp. 571–573. https://doi.org/10.1145/3626705.3631881
62. Wolling F, Pal G, Larcher L, Michahelles F. Smart trash can: Image classification and segmentation for waste quality management in the home. In: Proceedings of the 14th international conference on the Internet of Things. IoT '24. New York, NY, USA: Association for Computing Machinery; 2025. pp. 65–71. https://doi.org/10.1145/3703790.3703798
63. Longo E, Sahin FA, Redondi AEC, Bolzan P, Bianchini M, Maffei S. Take the trash out... to the edge. creating a smart waste bin based on 5G multi-access edge computing. In: Proceedings of the conference on information technology for social good. GoodIT '21. New York, NY, USA: Association for Computing Machinery; 2021. pp. 55–60. https://doi.org/10.1145/3462203.3475921
64. Aral RA, Keskin SR, Kaya M, Haclomeroglu M. Classification of TrashNet dataset based on deep learning models. In: IEEE international conference on big data (Big Data. 2018. pp. 2058–2062. https://doi.org/10.1109/BigData.2018.8622212

65. Fang B, Yu J, Chen Z, Osman AI, Farghali M, Ihara I, et al. Artificial intelligence for waste management in smart cities: a review. Environ Chem Lett. 2023;21(4):1959–89. https://doi.org/10.1007/s10311-023-01604-3.
66. Rovetta A, Xiumin F, Vicentini F, Minghua Z, Giusti A, Qichang H. Early detection and evaluation of waste through sensorized containers for a collection monitoring application. Waste Manage. 2009;29(12):2939–49. https://doi.org/10.1016/j.wasman.2009.08.016.
67. Huh J-H, Choi J-H, Seo K. Smart trash bin model design and future for smart city. Appl Sci. 2021;11(1111):4810. https://doi.org/10.3390/app11114810.
68. Tsita A, Eleftheriadis S, Vafeiadis T, Ioannidis D, Tzovaras D. AI-Assisted gamification platform for smart city food waste management. In: Companion proceedings of the 30th international conference on intelligent user interfaces. IUI '25 Companion. New York, NY, USA: Association for Computing Machinery; 2025. pp. 12–6. https://doi.org/10.1145/3708557.3716328.
69. Liu D, Zuo E, Wang D, He L, Dong L, Lu X. Deep learning in food image recognition: a comprehensive review. Appl Sci. 2025;15(14):7626. https://doi.org/10.3390/app15147626.
70. Bossard L, Guillaumin M, Van Gool L. Food-101 – Mining discriminative components with random forests. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) Computer Vision – ECCV 2014. Cham, Switzerland: Springer; 2014. pp. 446–461. https://doi.org/10.1007/978-3-319-10599-4_29
71. Kaur P, Sikka K, Wang W, Belongie S, Divakaran A. FoodX-251: A dataset for fine-grained food classification. 2019. arXiv:1907.06167, https://doi.org/10.48550/arXiv.1907.06167.
72. Min W, Liu L, Wang Z, Luo Z, Wei X, Wei X, Jiang S. Isia food-500: A dataset for large-scale food recognition via stacked global-local attention network. In: Proceedings of the 28th ACM International Conference on Multimedia. New York, NY, USA: Association for Computing Machinery; 2020. pp. 393–401. https://doi.org/10.1145/3394171.3414031
73. Marin J, Biswas A, Ofli F, Hynes N, Salvador A, Aytar Y, et al. Recipe1M+: A dataset for learning cross-modal embeddings for cooking recipes and food images. IEEE Trans Pattern Anal Mach Intell. 2021;43(1):187–203. https://doi.org/10.1109/TPAMI.2019.2927476.
74. Wu X, Fu X, Liu Y, Lim E-P, Hoi SCH, Sun Q. A large-scale benchmark for food image segmentation. In: Proceedings of the 29th ACM International Conference on Multimedia. MM '21. New York, NY, USA: Association for Computing Machinery; 2021. pp. 506–515. https://doi.org/10.1145/3474085.3475201
75. Okamoto K, Yanai K. UEC-FoodPix Complete: a large-scale food image segmentation dataset. In: Del Bimbo A, Cucchiara R, Sclaroff S, Farinella GM, Mei T, Bertini M, Escalante HJ, Vezzani R, editors. Lecture notes in computer science. Cham, Switzerland: Springer; 2021. vol. 12665, pp. 647–659. https://doi.org/10.1007/978-3-030-68821-9_51
76. Miller GA. Wordnet: a lexical database for English. Commun ACM. 1995;38(11):39–41. https://doi.org/10.1145/219717.219748.
77. Koskinopoulou M, Raptopoulos F, Papadopoulos G, Mavrakis N, Maniadakis M. Robotic waste sorting technology: toward a vision-based categorization system for the industrial robotic separation of recyclable waste. IEEE Robot Autom Magazine. 2021;28(2):50–60. https://doi.org/10.1109/MRA.2021.3066040.
78. Rother C, Kolmogorov V, Blake A. Grabcut: interactive foreground extraction using iterated graph cuts. ACM Trans Graph. 2004;23(3):309–14. https://doi.org/10.1145/1015706.1015720.
79. Ciocca G, Napoletano P, Schettini R. Food recognition and leftover estimation for daily diet monitoring. In: Murino V, Puppo E, Sona D, Cristani M, Sansone C, editors. New trends in image analysis and processing—ICIAP 2015 workshops. Cham, Switzerland: Springer; 2015. pp. 334–341. https://doi.org/10.1007/978-3-319-23222-5_41
80. Sahoo D, Hao W, Ke S, Xiongwei W, Le H, Achananuparp P, Lim E-P, Hoi SCH. Foodai: food image recognition via deep learning for smart food logging. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining. New York, NY, USA: Association for Computing Machinery; 2019. pp. 2260–2268. https://doi.org/10.1145/3292500.3330734
81. Xiao Z, Diao G, Deng Z. Fine grained food image recognition based on Swin transformer. J Food Eng. 2024;380:112134. https://doi.org/10.1016/j.jfoodeng.2024.112134.
82. Allegra D, Battiato S, Ortis A, Urso S, Polosa R. A review on food recognition technology for health applications. Health Psychol Res. 2020;8(3):9297. https://doi.org/10.4081/hpr.2020.9297.
83. Dalakleidi KV, Papadelli M, Kapolos I, Papadimitriou K. Applying image-based food-recognition systems on dietary assessment: a systematic review. Adv Nutr. 2022;13(6):2590–619. https://doi.org/10.1093/advances/nmac078.
84. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. Imagenet: a large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. 2009. pp. 248–255. https://doi.org/10.1109/CVPR.2009.5206848
85. Yanai K, Kawano Y. Food image recognition using deep convolutional network with pre-training and fine-tuning. In: 2015 IEEE International Conference on Multimedia and Expo Workshops (ICMEW). 2015. pp. 1–6. https://doi.org/10.1109/ICMEW.2015.7169816
86. Karabay A, Varol HA, Chan MY. Improved food image recognition by leveraging deep learning and data-driven methods with an application to Central Asian food scene. Sci Rep. 2025;15(1):14043. https://doi.org/10.1038/s41598-025-95770-9.

87. Chrintz-Gath G, Daivadanam M, Matta L, McKeever S. Image-based dietary assessment using the Swedish plate model: evaluation of deep learning-based You Only Look Once (YOLO) models. JMIR Formative Res. 2025;9:e70124. https://doi.org/10.2196/70124.
88. Ciocca G, Napoletano P, Schettini R. Food recognition: a new dataset, experiments, and results. IEEE J Biomed Health Inform. 2017;21(3):588–98. https://doi.org/10.1109/JBHI.2016.2636441.
89. Jocher G, Qiu J, Chaurasia A. Ultralytics YOLO. 2023. https://github.com/ultralytics/ultralytics
90. Lin T-Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P, Zitnick CL. Microsoft COCO: common objects in context. In: Computer vision–ECCV 2014: 13th European conference, Zurich, Switzerland, September 6-12, 2014, proceedings, Part V 13. 2014. pp. 740–755. https://doi.org/10.1007/978-3-319-10602-1_48. Springer
91. Ferreira J, Cerqueira P, Ribeiro J. Food waste detection in canteen plates using YOLOv11. Appl Sci. 2025;15(13):7137. https://doi.org/10.3390/app15137137.
92. Dhelia A, Chordia S. YOLO-based food damage detection: an automated approach for quality control in food industry. In: 8th international conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC). 2024. pp. 1444–1449. https://doi.org/10.1109/I-SMAC61858.2024.10714664
93. Liu H, Jiang Y, Zhang W, Li Y, Ma W. Intelligent electronic components waste detection in complex occlusion environments based on the focusing dynamic channel-you only look once model. J Clean Prod. 2025;486:144425. https://doi.org/10.1016/j.jclepro.2024.144425.
94. Ma W, Chen H, Zhang W, Huang H, Wu J, Peng X, et al. Dsyolo-trash: an attention mechanism-integrated and object tracking algorithm for solid waste detection. Waste Manage. 2024;178:46–56. https://doi.org/10.1016/j.wasman.2024.02.014.
95. Beery AJ, Eastman DW, Enos J, Richards W, Donnelly PJ. Smart compost bin for measurement of consumer food waste. In: Companion proceedings of the 26th international conference on multimodal interaction. ICMI '24 companion. New York, NY, USA: Association for Computing Machinery; 2024. p. 100–7. https://doi.org/10.1145/3686215.3686216.
96. Lo FPW, Sun Y, Qiu J, Lo B. Image-based food classification and volume estimation for dietary assessment: a review. IEEE J Biomed Health Inform. 2020;24(7):1926–39. https://doi.org/10.1109/JBHI.2020.2987943.
97. Luo Y, Kim Y-H, Lee B, Hassan N, Choe EK. Foodscrap: Promoting rich data capture and reflective food journaling through speech input. In: Proceedings of the 2021 ACM Designing Interactive Systems Conference. New York, NY, USA: Association for Computing Machinery; 2021. pp. 606–18. https://doi.org/10.1145/3461778.3462074
98. Radford A, Kim JW, Xu T, Brockman G, McLeavey C, Sutskever I. Robust speech recognition via large-scale weak supervision. In: Krause A, Brunskill E, Cho K, Engelhardt B, Sabato S, Scarlett J, editors. International conference on machine learning. Proceedings of machine learning research. 2023. vol. 202, pp. 28492–28518. https://proceedings.mlr.press/v202/radford23a.html
99. Zhang Y, Kang B, Hooi B, Yan S, Feng J. Deep long-tailed learning: a survey. IEEE Trans Pattern Anal Mach Intell. 2023;45(9):10795–816. https://doi.org/10.1109/TPAMI.2023.3268118.
100. Carion N, Gustafson L, Hu Y-T, Debnath S, Hu R, Suris D, Ryali C, Alwala KV, 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 HK, Dollár P, Ravi N, Saenko K, Zhang P, Feichtenhofer C. SAM 3: Segment Anything with Concepts. 2025. https://arxiv.org/abs/2511.16719
101. Vasu PKA, Pouransari H, Faghri F, Vemulapalli R, Tuzel O. MobileCLIP: fast image-text models through multi-modal reinforced training. In: IEEE/CVF conference on computer vision and pattern recognition (CVPR). 2024. pp. 15963–74. https://doi.org/10.1109/CVPR52733.2024.01511
102. Liang F, Wu B, Dai X, Li K, Zhao Y, Zhang H, Zhang P, Vajda P, Marculescu D. Open-vocabulary semantic segmentation with mask-adapted CLIP. In: Proceedings of the IEEE/cvf conference on computer vision and pattern recognition (CVPR). 2023. pp. 7061–70
103. Ding Z, Wang J, Tu Z. Open-Vocabulary Universal Image Segmentation with MaskCLIP. 2023. https://arxiv.org/abs/2208.08984
104. Xu M, Zhang Z, Wei F, Hu H, Bai X. Side adapter network for open-vocabulary semantic segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR). 2023. pp. 2945–54.
105. Wu X, Yu S, Lim E-P, Ngo C-W. OVFoodSeg: Elevating open-vocabulary food image segmentation via image-informed textual representation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR). 2024. pp. 4144–53.
106. Zhang Y, Unell A, Wang X, Ghosh D, Su Y, Schmidt L, Yeung-Levy S. Why are visually-grounded language models bad at image classification? In: Globerson A, Mackey L, Belgrave D, Fan A, Paquet U, Tomczak J, Zhang C, editors. Advances in neural information processing systems. 2024. vol. 37, pp. 51727–53. https://doi.org/10.52202/079017-1639
107. Cooper A, Kato K, Shih C-H, Yamane H, Vinken K, Takemoto K, et al. Rethinking VLMs and LLMs for image classification. Sci Rep. 2025;15(1):19692. https://doi.org/10.1038/s41598-025-04384-8.
108. Madan A, Peri N, Kong S, Ramanan D. Revisiting few-shot object detection with vision-language models. Adv Neural Inf Process Syst. 2024;37:19547–60. https://doi.org/10.52202/079017-0617

109. Beyer L, Steiner A, Pinto AS, Kolesnikov A, Wang X, Salz D, Neumann M, Alabdulmohsin I, Tschannen M, Bugliarello E, Unterthiner T, Keysers D, Koppula S, Liu F, Grycner A, Gritsenko A, Houlsby N, Kumar M, Rong K, Eisenschlos J, Kabra R, Bauer M, Bošnjak M, Chen X, Minderer M, Voigtlaender P, Bica I, Balazevic I, Puigcerver J, Papalampidi P, Henaff O, Xiong X, Soricut R, Harmsen J, Zhai X. PaliGemma: a versatile 3B VLM for transfer. 2024. https://doi.org/10.48550/arXiv.2407.07726

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Aidan Beery[1] · **Patrick J. Donnelly**[1]

✉ Aidan Beery
beerya@oregonstate.edu

Patrick J. Donnelly
donnellp@oregonstate.edu

[1] School of Electrical Engineering and Computer Science, Oregon State University, 25600 NW Monroe Ave., Corvallis, OR 97331, USA