IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 17 July 2024, accepted 2 September 2024, date of publication 9 September 2024, date of current version 20 September 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3456469

APPLIED RESEARCH

Contamination Detection From Highly Cluttered Waste Scenes Using Computer Vision

DISHANT MEWADA[1,2], CATHAOIR AGNEW[1,2], EOIN M. GRUA[1,2], CIARÁN EISING[1,2,3], (Senior Member, IEEE), PATRICK DENNY[1,4], (Member, IEEE), MARK HEFFERNAN[5], KEN TIERNEY[5], PEPIJN VAN DE VEN[1,2,3], AND ANTHONY SCANLAN[1,2,3]

[1]Data-Driven Computer Engineering (D[2]iCE) Group, University of Limerick, Limerick, V94 T9PX Ireland
[2]Department of Electronic and Computer Engineering, University of Limerick, Limerick, V94 T9PX Ireland
[3]CONFIRM Centre for Smart Manufacturing, University of Limerick, Limerick, V94 T9PX Ireland
[4]Department of Computer Science and Information Systems, University of Limerick, Limerick, V94 NX93 Ireland
[5]Advanced Manufacturing Control Systems Group, Limerick, V94 56R2 Ireland

Corresponding author: Dishant Mewada (dishant.mewada@ul.ie)

This work was supported by the Science Foundation Ireland (SFI) under Grant 16/RC/3918 (CONFIRM Centre).

ABSTRACT As the global production of waste continues to rise, there is a growing demand for more effective waste management strategies to handle this expanding problem. Recycling rates in the United States for recyclable materials are below 35%, resulting in elevated levels of waste being sent to landfills. This situation has alarming consequences, contributing to rising pollution in both soil and aquatic ecosystems, and is a significant source of concern for environmental scientists and the general population alike. The presence of contamination in recycling collection trucks is a root cause impacting recycling rates, leading to the rejection of entire loads from recycling processing sites. This problem can be alleviated by automatically detecting contamination in recyclable waste that is loaded into the collection truck hopper before compaction. In this paper, we have used different state-of-the-art computer vision-based models such as Faster-RCNN, Cascade-RCNN, Retinanet, YOLOv8 and Mask-RCNN to identify contamination within a densely cluttered waste environment. We further investigate the viability of transfer learning, comparing it to the models trained from scratch. The YOLOv8-x model attained a mean-average precision of 0.395 without using transfer learning, whereas with the incorporation of transfer learning, its performance increased to 0.463.

INDEX TERMS Transfer learning, computer vision, waste management, instance segmentation, object detection, supervised learning, neural network.

I. INTRODUCTION

In the United States, approximately 94 million tons (measured in U.S. short tons unless stated otherwise) of Municipal Solid Waste (MSW) underwent recycling and composting processes, resulting in a composting and recycling rate of about 32.1%. In 2018, the United States generated an immense 292.4 million tons of MSW, translating to an average daily waste production of 4.9 pounds per person. Out of the overall MSW produced, around 69 million tons underwent recycling, while 25 million tons were subjected to composting [1].

The management of such a large volume of waste is becoming a more critical issue as a result of the growing global population and the increase in waste production. Waste production is predicted to rise from its present level of approximately 2.1 billion tonnes annually to 2.6 billion tonnes by the year 2030 [26]. The rise in waste generation presents various difficulties, such as environmental contamination, depletion of resources, and potential threats to public health. To mitigate the damaging environmental consequences of increased waste production, effective recycling solutions are essential.

The associate editor coordinating the review of this manuscript and approving it for publication was Claudio Loconsole.

Recycling contamination is a global issue. For many years, China handled a huge portion of the world's recycling. However, starting with Operation Green Fence [13] in 2013, the Chinese authorities aimed to significantly curb the amount of contaminated waste that it imported. This was made even more aggressive in 2017 when the Operation Sword policy [49] was implemented by the Chinese authorities, directly banning the import of 24 material types, and effectively banning even more through extremely stringent contamination requirements. This sequence of steps to curtail recycling imports into China has had a large effect on the global waste trade [8].

Other countries are following suit, particularly Southeast Asian countries, with contamination resulting in a refusal to accept waste [28]. These developments have provided significant impetus to better recyclable waste contamination management at source. In general, contamination is identified at processing plants through manual or automatic processes. Automated processes face a distinctive challenge due to the following factors: a significant amount of clutter, the existence of objects that are both highly deformable and translucent, and subtle distinctions between different object classes [4].

Automating waste sorting requires machines with the capability to visually identify and manage a wide range of items. The complexity arises from the endless variety of objects in terms of size, shape, color, brightness, and conditions, such as objects being dirty, broken, crushed, or overlapped [2]. These challenges make it difficult to identify and separate waste types. As sorting centers receive large volumes of waste from garbage trucks, it is currently not possible to identify individual culprits for the contamination. Whilst it is possible to establish that the waste load of a given truck is contaminated, it is currently not possible to establish from which individual source the contamination originates.

We addressed this problem within our project by implementing a system that detects the recycling waste contamination at the pickup. Each garbage truck is equipped with a camera, which records video footage during pickup. We extract suitable images from this footage that contain the targeted contaminated items. These images are used to train the models and later to detect the contamination. By using this approach we can identify which area is generating more contaminants in their recycling wastes.

The main contribution of this study is the investigation of the identification of recyclable waste contamination at the source, during pick up, in an industrial real-world setting. Identification of contaminants in collected bins is important to allow the waste collection company or other parties involved to engage with non-compliant businesses or individuals. Cutting-edge computer vision models, commonly referred to as state-of-the-art (SOTA), are employed to detect contaminants in recycled waste. One of the key advantages of computer vision technology is its cost-effectiveness, as a single camera can be installed on each truck in a fleet to achieve desired outcomes.

The paper follows a structured approach, starting with a review of previous studies in Section II. Section III talks about the overview of sota models and backbones used in the paper. Section IV delves into the details of the experimental setup and methodology. Section V presents the findings of the experiment, followed by an in-depth analysis and discussion in Section VI. Finally, Section VII condenses the key takeaways and conclusions extracted from the research.

II. RELATED WORK

In the past, numerous research projects have focused on using neural networks and machine learning for waste detection and classification in images. Nevertheless, none of these studies specifically address the challenge of detecting waste within real-world cluttered environments.

A detailed comparison of related work is shown in Table 1, which contains the method used in the published experiments, a description of the dataset, and the application field of the research. We further focus on whether the dataset contains a cluttered background and if the images were captured in a real-world setting.

Yang and Thung worked on classifying recycling waste using support vector machines (SVMs) [50]. Their dataset consisted of 2,527 images and included six categories of waste: glass, paper, cardboard, metal, plastic, and trash. Each image is labeled with its corresponding category. It primarily consists of photographs of various types of waste captured against a white background. Each photo was taken with a unique combination of exposure and lighting, focusing solely on highlighting a single object. The authors have explored the application of SVM and Convolutional Neural Network (CNN) algorithms for efficiently classifying waste items into six specific recycling categories. They adopted a modified architecture inspired by AlexNet. In their experimentation, the SVM outperformed the CNN, attaining a 63% accuracy. In contrast, the CNN only managed to achieve a testing accuracy of 27%.

Another major contribution in the field of litter detection was made by Proença and Simões where 1500 images were captured and labelled into 28 categories and named the TACO dataset. They used a Mask R-CNN model to detect waste such as cigarettes, bottles, and plastic bags [38]. In the same year, Panwar et al. compared one-stage and two-stage detectors for detecting marine debris. They experimented with Trashnet and manually annotated the TACO dataset. One drawback of the combined dataset was its absence of objects set against cluttered backgrounds [37].

Bobulski and Kubanek explored plastic waste classification using approximately 100k images. They used a conveyor belt, air jet, processing unit, and a camera in their proposed system for plastic waste sorting [6]. They achieved 74% accuracy in detecting plastic waste in 4 different categories. The dataset lacked a real-world setting and a cluttered background.

In 2022, Bashkirova et al. worked on recycling waste using object detection and instance segmentation. The images were collected by rotating a conveyor belt and also placing waste material on it, which was highly cluttered [4]. The sole drawback of the dataset is its lack of resemblance to a real-world scenario. This study aligns most closely with our own research. They used different supervised methods to detect 20 types of recycling waste from the images. While their ZeroWaste dataset accurately reflects the cluttered environment seen in a real-life setting, the controlled lighting of this environment is less challenging for computer vision models, which is still a limitation of this dataset. Real-world applications may not have the means to implement consistent lighting as used in the ZeroWaste dataset. Occlusions created by shadows or artifacts such as snow, along with the intensity of light varying by location, seasons, and the time of the day, maybe influential factors for real-world scenarios that may impact performance. The authors provided baseline results for popular fully, semi-supervised, weakly supervised, and transfer learning techniques and concluded that existing SOTA detection and segmentation algorithms are incapable of handling this complicated real-world setting effectively.

Kumsetty et al. presented a novel approach leveraging quantum transfer models for waste classification. Researchers explored transfer learning techniques by utilizing pre-trained models on extensive datasets like ImageNet and COCO [27]. They introduced TrashBox, a comprehensive waste dataset encompassing 17,785 images categorized into seven classes: cardboard, e-waste, glass, metal, medical waste, paper, and plastic. However, this dataset lacked real-world representation. Using a fine-tuned ResNet-101 model resulted in an impressive accuracy of 98.47%.

Majchrowska et al. used a novel two-stage detection methodology to detect household litter [33]. In stage 1, they detected litter with an object detector. The detected region is then cropped and used for the classification. The dataset encompasses seven categories of waste: biodegradable, glass, metals, plastics, non-recyclable, paper, and miscellaneous items (construction and demolition debris, large waste such as tires, electronics, and appliances, batteries, paint, and varnish containers, or expired medication), as well as unidentified litter that is highly decomposed or difficult to distinguish. They achieved 75% accuracy in classifying waste, which was mainly found in the surrounding area of the household.

A. YOLO IN WASTE DETECTION

Fulton et al. investigated detecting debris (different plastic materials) in marine environments [14]. The images captured for the dataset were resembling the real-world setting of the underwater environment but without cluttered backgrounds. The project employed four network architectures chosen from widely recognized and effective object detection networks. These architectures encompassed YOLOv2, Tiny-YOLO, SSD with MobileNet v2, and Faster RCNN with Inception v2. Another similar work was undertaken by Hong et al. where they used object detection and segmentation on their Trashcan dataset which they created using 7,212 images [22]. The dataset included observations of waste, as well as a diverse range of underwater plants and animals captured by remotely operated vehicles (ROVs). The authors employed Faster R-CNN with a ResNeXt-101-FPN backbone, as well as Mask R-CNN with an X-101-FPN backbone.

In a recent research study on waste detection, Mao et al. employed the YOLOv3 algorithm to identify six distinct waste categories: plastic containers, bottles, metals, cartons, paper containers, and glass. Their YOLOv3 model achieved mAP@0.5 (mean average precision) [23] of 92.12%, based on 3591 training images. Although the model achieved a high mean average precision score, its performance may not generalize well to real-world scenarios due to the limited scope of its training data. The dataset used for training did not include images with cluttered backgrounds or real-world settings [34].

In the year 2023, Demetriou et al. utilized single-stage and two-stage detection models for real-time object detection tasks on the Construction and Demolition Waste (CDW) dataset, which contains 6600 images with three categories: brick, concrete, and tile [12]. Their assembly consisted of an industrial camera overlooking the materials on a conveyor belt and robotic manipulators for sorting waste material. They used models such as Faster-RCNN, SSD and YOLO with different backbones to detect the materials, with YOLOv7 achieving the highest mAP of 71.7%.

In the recent work Sarswat et al., the authors of the paper investigated the performance of YOLOv5 and YOLOv7 models for e-waste component detection using a dataset comprising images of various e-waste classes such as copper, PCB, steel, glass, and aluminum [43]. Their study tested multiple models under different conditions, focusing on variations in batch size, epochs, and input image size. YOLOv7 showed optimal performance with a prediction accuracy of approximately 94%, achieving high F1 scores of 1.0 for copper, PCB, and plastic and mAP values exceeding 0.96 for these classes. In contrast, YOLOv5 demonstrated the highest precision of 0.98 for PCB but generally performed less effectively compared to YOLOv7. The research highlighted the superior detection capabilities of YOLOv7 for complex e-waste classification tasks.

The paper [36] explores the application of deep learning techniques to enhance the efficiency of waste sorting processes. The authors introduce the WaRP (Waste Recycling Plant) dataset, which includes 28 categories of recyclable waste, for training various models. The dataset contains 2452 training images and 522 test images. The authors highlight the challenges posed by class imbalance in the dataset. For classification tasks, models such as CNN, LeNet-5, AlexNet, VGG16, MobileNet-v2, Inception-v3, and DenseNet are employed to categorize waste items. For detection tasks, the YOLOv8 model is used to identify waste in complex and cluttered scenes.

TABLE 1. Comparison of the related work: • Real-world setting: The data was collected in a real-world setting. –✓ Data collected from real-world settings, in natural, non-simulated form, surrounding was not controlled by any means –*X* Simplified data • Densely Cluttered Environment: This implies the data was collected in a densely cluttered scene, this may be multiple instances of the same object (e.g., multiple glass bottles) or multiple instances of different objects with significant overlapping or occlusion apparent. –✓ Densely cluttered data collected –*X* Non-cluttered/plain background scenes.

Related Work	Vision Task	Number of images	Number of categories	Densely Cluttered Environment	Real-world Setting	Method used	Application
[50]	Classification	3000	6	×	×	SVM	Recycling waste
[14]	Object Detection	5720	3	×	✓	YOLOv2 Tiny-YOLO Faster R-CNN SSD	Marine debris
[22]	Object Detection Inst. Seg.	7212	22	×	✓	Faster R-CNN Mask R-CNN	Marine debris
[37]	Object Detection	369	4	×	✓	Faster R-CNN RetinaNet	Marine debris
[38]	Inst. Seg.	1500	28	×	✓	Mask R-CNN	Litter detection
[6]	Classification	109820	4	×	×	AlexNet MobineNetV1 MobileNetV2	Plastic Waste
[4]	Object Detection Inst. Seg.	10715	20	✓	×	RetinaNet Mask R-CNN TridentNet supervised learning	Recycling waste
[27]	Classification	17785	7 (25)	×	×	ResNet-34 ResNet-50 ResNet-101 DenseNet-121 VGG-19	Trash Objects
[33]	Object Detection + Classification	28000 app.	7	×	✓	EfficientDet DETR Mask RCNN	waste detection
[34]	Object Detection	6233	6	×	×	YOLOv3	waste detection
[12]	Object Detection	6600	3	✓	×	Faster R-CNN SSD YOLOv5,v6,v7	Construction Material sorting
[43]	Object Detection	400	6	×	×	YOLOv5 YOLOv7	waste detection
[36]	Classification Object Detection	2974	28	✓	×	LeNet-5 AlexNet VGG16 MobileNet-v2 Inception-v3 DenseNet YOLOv7	waste detection
This Work	Object Detection Inst. Seg.	3591	6	✓	✓	Faster R-CNN Mask R-CNN RetinaNet Cascade RCNN YOLOv8	Contamination from Waste

III. METHODOLOGY

The purpose of the study is to apply and assess various state-of-the-art computer vision models for automatically detecting contamination in recyclable waste within real-world, cluttered environments. By leveraging models like Faster-RCNN, Cascade-RCNN, Retinanet, YOLOv8, and Mask-RCNN and using a unique, annotated dataset (Contamination Dataset) reflective of actual waste collection scenarios, our

study seeks to improve the efficiency and accuracy of contamination detection. The study particularly emphasizes the use of transfer learning, demonstrating its advantage over models trained from scratch.

In this section, we discuss the various methods employed in our study, providing an overview of the state-of-the-art (SOTA) computer vision models used with their respective backbones. We further discuss various techniques used for model training, such as transfer learning and training from scratch.

A. OVERVIEW OF SOTA MODELS USED

1) FASTER R-CNN

Faster R-CNN [42], introduced in 2015, is a state-of-the-art two-stage object detection framework which is an improvement over earlier algorithms (R-CNN [17], SPPNet [18], and Fast R-CNN [16]) that relied on the slow selective search method to identify region proposals. Faster R-CNN introduces a Region Proposal Network (RPN) to generate region proposals directly from the image features calculated by the CNN during the classification step. The RPN uses a 3×3 sliding window across CNN feature maps to generate multiple region proposals with various shapes and sizes. These proposals are passed to the RoI pooling layer, which extracts fixed-size feature maps for each proposal. These feature maps are then classified, and the bounding boxes are predicted.

2) MASK R-CNN

Mask R-CNN [20] extends Faster R-CNN to address instance segmentation, combining object detection and semantic segmentation. Unlike Faster R-CNN, which outputs class labels and bounding box offsets, Mask R-CNN adds an object detection mask for each proposal. It uses the RoIAlign layer to precisely map features to input positions, addressing misalignment issues in RoI pooling. Mask R-CNN has achieved state-of-the-art performance in instance segmentation.

3) CASCADE R-CNN

Cascade R-CNN [9] is an object detection framework that improves performance by tackling overfitting and inference-time mismatches at higher Intersections over Union (IoU) thresholds. It uses a multi-stage approach, where each detector in the sequence is trained with progressively higher IoU thresholds, increasing selectivity against close false positives. This sequential training, where each stage uses the output of the previous stage, ensures a sufficient number of positive samples and reduces overfitting. During inference, the same cascade structure is used, aligning hypotheses closely with the detector quality at each stage, resulting in better detection performance.

4) RETINANET

RetinaNet [31] is a single-stage object detector similar to SSD [32] and YOLO [41], but it achieves performance comparable to two-stage detectors like Faster R-CNN. It introduces a novel focal loss function that addresses class imbalance during training by down-weighting well-classified examples, thus focusing on harder examples. RetinaNet consisted of a Feature Pyramid Network backbone, based on ResNet50 or ResNet101, and two task-specific subnetworks for classification and bounding box regression. The classification subnetwork uses a series of convolutional layers and applies focal loss, while the regression subnetwork has a similar structure without shared parameters.

5) YOLO

YOLO (You Only Look Once), is a series of real-time single-stage object detection models that have evolved significantly since its inception. YOLOv1, released in 2015, [41] treats object detection as a regression problem, predicting class probabilities and bounding boxes from entire images in a single evaluation. This approach integrates the detection pipeline into a single neural network, allowing for end-to-end optimization directly on detection performance. The network utilizes features from the entire image to predict bounding boxes and their associated classes simultaneously, enabling it to consider the global context of the image and all contained objects for more accurate predictions. YOLOv2 [39], introduced in 2017 as YOLO9000, brought major improvements with batch normalization, fine-tuning using larger images, and the use of anchor boxes, achieving better recall and mAP through k-means clustering. YOLOv3 [40] further enhanced the model by implementing a logistic classifier for multi-label classification, predicting objects at three scales, and employing a more complex loss function, resulting in robust detection performance. YOLOv4 [7] utilizes techniques such as Spatial Pyramid Processing (SPP) to extract features from the image at different resolutions & scales and Cross Stage Partial connection (CSP) is used to improve the model's accuracy. YOLOv5 [24], released in 2020, focuses on balancing inference speed and accuracy, offering various model sizes and incorporating CSPDarknet, PANet, and advanced training techniques. YOLOv6 [29] distinguishes itself from previous versions by employing a more lightweight and efficient neural network architecture based on the EfficientNet-Lite family, enabling faster performance with fewer computational resources. It also enhances robustness and generalization through data augmentation techniques such as rotation, scaling, and flipping to input images during training. YOLOv7 [47], one of the recent iterations, surpasses previous versions and other models in speed and accuracy, utilizing architectural innovations and compound scaling without relying on pre-trained backbones. YOLOv8 [25] is a state-of-the-art model that substantially improves object detection, classification, and segmentation tasks. YOLOv8 introduces several key enhancements, including an anchor-free detection system, improved convolutional blocks, and mosaic augmentation techniques applied during training. Additionally, the model

incorporates a new C2f (Coordinates-To-Features) module in its backbone to transform coordinate information into feature representations efficiently, a decoupled head to independently handle classification and regression tasks, and a refined loss function with task alignment scores. We have used the latest YOLOv8 model in our work. YOLOv8 offers several variants tailored for different applications and computational requirements, such as YOLOv8-n (Nano), YOLOv8-s (Small), YOLOv8-m (Medium), YOLOv8-l (Large), YOLOv8-x (Extra Large). YOLOv9, launched in February 2024 [48], incorporates the Programmable Gradient Information (PGI) framework and the Generalized Efficient Layer Aggregation Network (GELAN). These enhancements resolve information bottlenecks in deep neural networks, improve accuracy for lightweight architectures, and enhance object detection performance, making the model efficient and versatile for various devices, including those with limited resources [3]. Released in the same year, YOLOv10 [45] advances real-time object detection for uses like agricultural monitoring and autonomous driving by eliminating the need for non-maximum suppression (NMS) during post-processing, thus enhancing inference speed. Its architectural improvements reduce computational demands, and extensive testing shows it improves latency and model size while maintaining high detection accuracy, especially on the COCO dataset. We have opted to use the YOLOv8 model because YOLOv9 and YOLOv10 are still in the early stages of widespread adoption and achieving implementation stability.

B. OVERVIEW OF BACKBONES USED

Resnet: Introduced by [19] in 2015, the key innovation of ResNet (Residual Network) is the use of residual connections, which are shortcut paths that skip one or more layers. These connections address the vanishing gradient problem, making it possible to train networks with hundreds or even thousands of layers. It is possible to construct complex neural network models due to Resnet's effective feature extraction and enhanced gradient flow during training. ResNet backbones, including ResNet-50, ResNet-101, and ResNet-152, are extensively used in computer vision applications such as image classification, object detection, and segmentation. In this paper, we have utilized the ResNet backbone containing 101 layers for the models Faster R-CNN, Cascade R-CNN, Mask R-CNN, and Retinanet.

1) CSPDARKNET

CSPDarknet is a variant of the Darknet architecture, used as the backbone for YOLOv4 [7]. It builds on DarkNet-53 and employs a Cross-Stage Partial Network (CSPNet) [46] strategy to split the feature map of the base layer into two parts, which are then merged through a cross-stage hierarchy. This split and merge approach enhances gradient flow throughout the network. In this paper, we have used YOLOv8, which uses a custom CSPDarknet53 backbone.

C. VARIOUS TECHNIQUES EMPLOYED FOR MODEL TRAINING

1) TRAINING FROM SCRATCH

Training from scratch involves initializing the weights of a neural network randomly and then training the network on the target task from the beginning. This process can be time-consuming and requires a large amount of labeled data to achieve high performance because the models need to learn all the features from the new data without any prior knowledge [44].

2) TRANSFER LEARNING

Transfer learning, in contrast, involves using ''knowledge gained'' while solving one problem and applying it to a different but related problem. The goal is to utilize a pre-trained model for a new task. This method begins with a pre-initialized weight matrix obtained from training on a source task, which helps the model to learn more efficiently when applied to a new target task [44]. Using transfer learning, pre-trained networks can be applied to smaller datasets by simply fine-tuning the final fully connected layers of the CNN [21].

IV. EXPERIMENTAL SETUP

In this section, we provide an overview of our experimental setup. In the Contamination Dataset section, we detail the positioning of the camera & its calibration, dataset distribution, and the distinctive real-world characteristics of our dataset. In the ''Training Configuration'' section, we outline the system specifications, the rationale for model selection, the implementation & training processes, and various hyper-parameters applied during training.

A. CONTAMINATION DATASET

This study relied on proprietary data (Contamination dataset) collected from a camera fitted with a 1/2.9-inch 2-megapixel CMOS color image sensor, capturing frames at a rate of ten frames per second and with an electronic shutter speed of 1/30s. The sensor was able to operate in a range of temperatures, from -20°C to 70°C.

Fig. 2 shows the position of the camera inside the truck, while Fig. 3 illustrates the front view of the camera. Video records were compiled from collection drives spanning various regions in the US and EU. Contamination-containing frames were carefully chosen from the video and labeled for use in developing the models. All the images were at a resolution of 704 × 480 pixels with no post-processing and modifications. Fig. 1 illustrates a typical instance of an annotated image sourced from the Contamination dataset.

The number of images and object classes in each dataset is summarized in Table 2. Table 3 presents a quantitative breakdown of the proportion of *Small*, *Medium*, and *Large* in each class for the training, validation, and testing datasets. COCO's standard definitions of object size [23] are used for this categorization.

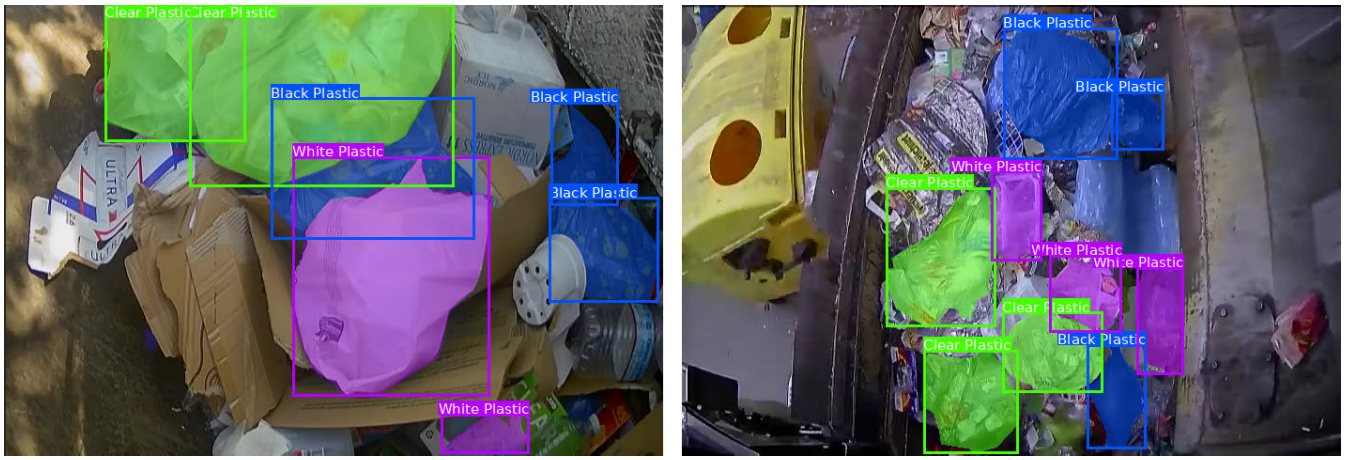

FIGURE 1. Annotation example for the contamination dataset.

FIGURE 2. Position of the camera inside the garbage truck.

FIGURE 3. A frontal perspective of the camera positioned within the truck.

White plastic and black plastic are the classes that contain the major instances found in the dataset, with 7,579 and 1,916 instances, respectively. A train/test/val split of 55.7% (1999 images) / 14.6% (522 images) / 29.8% (1068 images) is used.

TABLE 2. Dataset distribution.

No. of Instances	Black Plastic	White Plastic	Clear Plastic	Polystyrene	Blue Bag
Train	1,058	4,216	456	564	394
Val	296	1,106	108	156	84
Test	562	2,257	244	272	261
Total	1,916	7,579	808	995	739

The uniqueness of the Contamination dataset is characterized by its real-world attributes. Since the images are sourced from both commercial and residential settings, the training data accurately represents the operational conditions necessary for a robust computer vision system designed to detect contamination. The dataset's real-world setting introduces several challenges, such as a cluttered background, diverse object distances, and fluctuating lighting and weather conditions. These variations can impact the distribution of pixels and potentially hinder a model's performance. The interplay of these factors poses a significant hurdle for computer vision applications.

B. TRAINING CONFIGURATION

We employed the MMdetection toolkit (version 2.28.2) [10] as well as MMYOLO toolkit (version 0.6.0) [11] in conjunction with CUDA 11.8 [35] to train a comprehensive range of cutting-edge object detection and instance segmentation models. Weights & Biases (WandB) [5] tool is used to track the model results and real-time performance of the models. The experiments were executed on a specialized workstation powered by four NVIDIA A100-SXM4-40GB GPUs.

FIGURE 4. Example Images from the Contamination Datasets.

TABLE 3. Breakdown of the dataset.

Classes	No. of Instances [%]	Small [%]	Medium [%]	Large [%]
Train				
Black Plastic	1058 (15.8)	0 (0.0)	336 (31.8)	722 (68.2)
White Plastic	4216 (63.0)	8 (0.2)	2340 (55.5)	1868 (44.3)
Clear Plastic	456 (6.8)	0 (0.0)	189 (41.4)	267 (58.6)
Polystyrene	564 (8.5)	0 (0.0)	279 (49.5)	285 (50.5)
Blue Bag	394 (5.9)	0 (0.0)	183 (46.4)	211 (53.6)
Validation				
Black Plastic	296 (16.9)	0 (0.0)	93 (31.4)	203 (68.6)
White Plastic	1106 (63.2)	0 (0.0)	609 (55.1)	497 (44.9)
Clear Plastic	108 (6.2)	0 (0.0)	36 (33.0)	72 (66.7)
Polystyrene	156 (8.9)	0 (0.0)	75 (48.1)	81 (51.9)
Blue Bag	84 (4.8)	0 (0.0)	39 (46.4)	45 (53.6)
Test				
Black Plastic	562 (15.6)	0(0.0)	184 (32.7)	378 (67.3)
White Plastic	2257 (62.8)	7 (0.3)	1285 (56.9)	965 (42.8)
Clear Plastic	244 (6.8)	0 (0.0)	93 (38.1)	151 (61.9)
Polystyrene	272 (7.6)	0 (0.0)	133 (48.9)	139 (51.1)
Blue Bag	261 (7.2)	2 (0.7)	133 (51.0)	126 (48.3)

The experiments utilized MMdetection toolkit for training Faster-RCNN [42], RetinaNet [31], Cascade-RCNN [9] and MMYOLO toolkit for training YOLOv8-x [25] for object detection task, while Mask-RCNN [20] was employed for instance segmentation.

Models were first trained on the COCO dataset [30], and the model weights were used for transfer learning on the Contamination dataset. Faster-RCNN, RetinaNet, Cascade-RCNN, and Mask-RCNN were trained with ResNet-101 backbone [19] while YOLOv8-x was trained with the backbone CSPDarknet [7].

In our study, we selected the models Faster R-CNN, Cascade R-CNN, RetinaNet, Mask R-CNN, and YOLOv8-X. These models were chosen for their similarity in the number of parameters, which ensures a more equitable comparison of their performance metrics. The table 4 shows the number of parameters used by different models used in the experimentation.

TABLE 4. Number of parameters comparison.

Models	Faster-RCNN	Cascade-RCNN	Retinanet	Mask-RCNN	YOLOv8-x
number of parameters (in Millions)	60.12	87.93	55.18	62.76	68.16

For models Faster-RCNN, RetinaNet, Cascade-RCNN, and Mask-RCNN, images were processed in batches of 2. The training process employed the following parameters for the models:

- Number of epochs: 24
- Optimizer: Stochastic Gradient Descent (SGD)
- Learning rate: 0.02
- Momentum: 0.9
- Weight decay: 0.0001

Furthermore, for these models, a learning rate scheduler was incorporated, leading to a 10-fold reduction in the learning rate at epochs 16 and 22.

For the YOLOv8-x model, following parameters were used in MMYOLO toolkit:

- Number of epochs: 500 with early stopping
- Batch size: 16
- Optimizer: SGD
- Learning rate: 0.005
- Momentum: 0.9
- Weight decay: 0.0005

For comparison, we trained models on the Contamination dataset with and without using COCO for pre-trained weights, as shown in the Tables 7, 8, 9, and 10.

V. RESULTS

A test set containing 1,068 images was used for inference. For our experiments, we have primarily focused on COCO's main evaluation metric, $mAP_{0.50:0.05:0.95}$ [23].

TABLE 5. Object Detection Results on COCO dataset (From Scratch).

Model	mAP	mAP_50	mAP_75	mAP_S	mAP_M	mAP_L
Faster-RCNN	0.405	0.612	0.441	0.223	0.437	0.516
Cascade-RCNN	0.429	0.614	0.468	0.237	0.460	0.547
Retinanet	0.396	0.591	0.424	0.213	0.426	0.511
Mask-RCNN	0.412	0.616	0.450	0.227	0.442	0.526
YOLOv8-x	0.528	0.701	0.574	0.334	0.571	0.671

TABLE 6. Instance Segmentation Results on COCO dataset (From Scratch).

Model	mAP	mAP_50	mAP_75	mAP_S	mAP_M	mAP_L
Mask-RCNN	0.371	0.588	0.398	0.192	0.398	0.490

TABLE 7. Object Detection Results on Contamination dataset (Transfer Learning).

Model	mAP	mAP_50	mAP_75	mAP_S	mAP_M	mAP_L
Faster-RCNN	0.408	0.600	0.439	0.050	0.332	0.474
Cascade-RCNN	0.423	0.591	0.462	0.103	0.338	0.500
Retinanet	0.442	0.642	0.480	0.074	0.370	0.507
Mask-RCNN	0.423	0.604	0.461	0.113	0.342	0.496
YOLOv8-x	0.463	0.605	0.500	0.376	0.380	0.538

TABLE 8. Instance Segmentation Results on Contamination dataset (Transfer Learning).

Model	mAP	mAP_50	mAP_75	mAP_S	mAP_M	mAP_L
Mask-RCNN	0.424	0.598	0.462	0.033	0.309	0.536

TABLE 9. Object Detection Results on Contamination dataset (From Scratch).

Model	mAP	mAP_50	mAP_75	mAP_S	mAP_M	mAP_L
Faster-RCNN	0.372	0.595	0.403	0.069	0.313	0.429
Cascade-RCNN	0.383	0.585	0.410	0.072	0.318	0.444
Retinanet	0.359	0.577	0.385	0.230	0.292	0.416
Mask-RCNN	0.384	0.613	0.407	0.082	0.323	0.442
YOLOv8-x	0.395	0.571	0.420	0.310	0.313	0.468

A. STATE-OF-THE-ART RESULTS

Detection models were first trained from scratch on the COCO dataset so that the respective trained model weights could be used for transfer learning on the Contamination dataset. Tables 5 and 6 show the object detection and instance segmentation results for the models trained on the COCO dataset.

Further, in tables 7 and 8, results for the models trained using transfer learning are shown. In this case, the model weights trained on the COCO dataset are utilized for transfer learning on the Contamination dataset.

In tables 9 and 10, results for the models trained on the Contamination dataset with randomly initialized weights are shown.

B. EXPLORATORY ANALYSIS OF INFERENCE RESULTS

Transfer learning shows an improvement in mAP compared to the model trained from scratch on the Contamination dataset. Table 11 shows the overall mAP difference after applying transfer learning compared to the models trained with randomly initialized weights.

TABLE 10. Instance Segmentation Results on Contamination dataset (From Scratch).

Model	mAP	mAP_50	mAP_75	mAP_S	mAP_M	mAP_L
Mask-RCNN	0.396	0.604	0.444	0.008	0.284	0.505

TABLE 11. Overall mAP increase using transfer learning.

Models	Faster-RCNN	Cascade-RCNN	Retinanet	Mask-RCNN (bbox)	Mask-RCNN (segm)	YOLOv8-x
mAP increase	0.036	0.040	0.083	0.039	0.028	0.068
% mAP increase	9.68	10.44	23.12	10.16	7.07	17.21

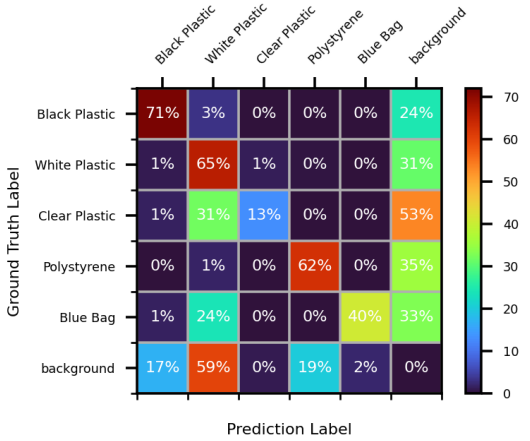

FIGURE 5. Normalized Confusion Matrix (Model: YOLOv8-x using transfer learning).

By examining the model's predictions for a sample of images from the Contamination dataset (Fig. 6), we can gain valuable insights into its performance. By analyzing both its successes and failures, we can better understand its strengths and weaknesses. For a more comprehensive assessment, we can also review the confusion matrix for the entire test dataset (Fig. 5). To account for false positives and false negatives associated with the targeted classes, MMDetection utilizes an additional background category in its confusion matrix. MMDetection's confusion matrix includes a separate class for background objects to track false positives and negatives related to the classes of interest. The confusion matrix calculation does not account for accurately identifying the background class, as it is not considered one of the classes of interest.

From the confusion matrix in Fig. 5, we observe that the model performs best with Black Plastic (71%), White Plastic (65%), and Polystyrene (62%), although some misclassification into Background still occurs. Clear Plastic (13%) and Blue Bag (40%) show significant misclassification with other categories, particularly Background and White Plastic, due to

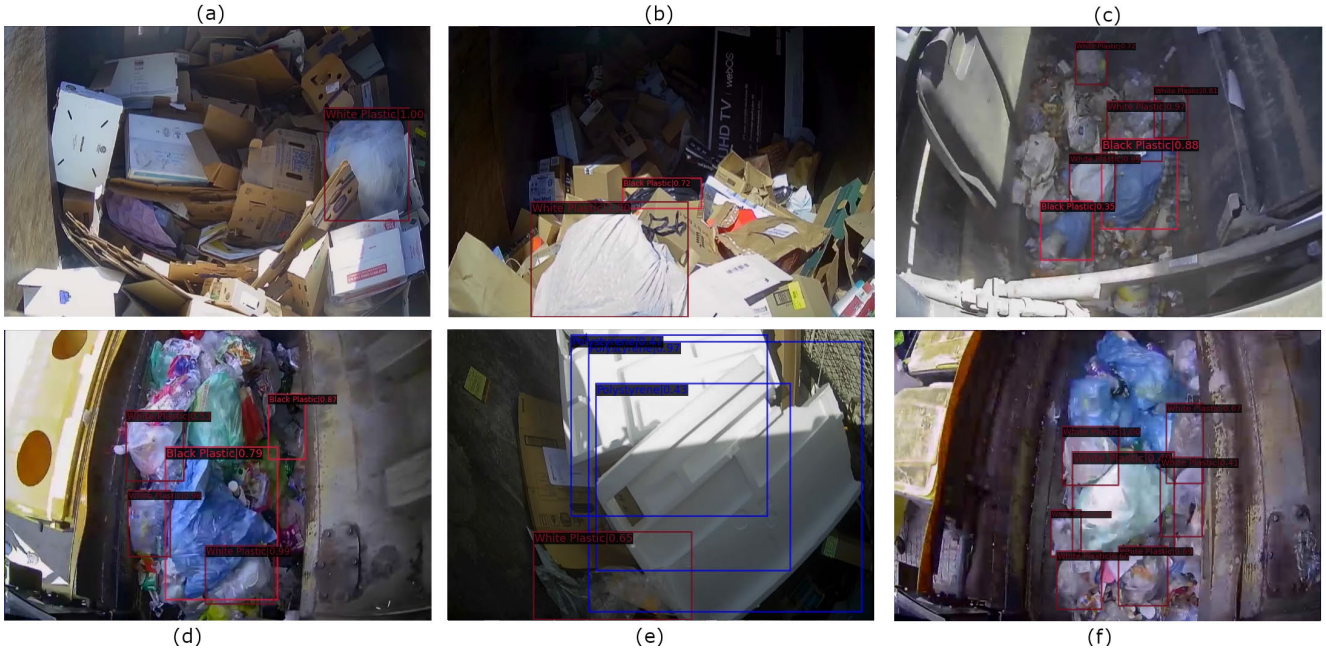

FIGURE 6. Inference results on the Contamination test dataset (model: Retinanet). In examples (a), (b) there are correct identification even when the objects were in different environments with varying backgrounds and lighting conditions. In examples (c), (d) there are incorrect predictions of the blue bag as black plastic. In example (e) there is multiple detection of polystyrene. In example (f) there is missing detection for the blue bag.

similar visual features, resulting in lower accuracy for these classes.

VI. DISCUSSION

The proprietary Contamination dataset has allowed exploration of the feasibility of recycling waste contamination detection at the source. Since the data were gathered from a real-world commercial and domestic collection route, they accurately represent the necessary real-world conditions for detecting contaminants from recyclable materials, as the images represent a variety of light and weather conditions and a densely cluttered background. The transfer learning score displayed in Tables 7 and 8 outperforms the model trained from scratch (Tables 9 and 10) in terms of mAP. While the acceptable tolerance rate may vary depending on the specific use case, these results demonstrate the potential for integrating computer vision as a contamination detection tool.

Table 11 shows a trend of better performance for each model trained using COCO pre-trained weights. The transfer learning yields better results than the models trained from scratch on the Contamination dataset. This demonstrates the effectiveness of pre-training with the COCO dataset for this application, which only has 1999 images in the training dataset. The YOLOv8-x model outperformed other models with 0.463 mAP. The mAP score for the YOLOv8-x model increased from 0.395 to 0.463, whereas Retinanet had the highest increase in mAP when using pre-training with COCO versus training from scratch.

The dataset has some deficiencies that can be addressed by collecting more data in the future. For instance, there are fewer instances of small area objects present in the dataset (Table 2), which is likely the reason why the mAP for large and medium area bounding boxes is higher compared to the smaller area ones. Increasing small-area objects can yield better overall mAP results. The Contamination dataset has more instances of white plastic and black plastic compared to the other categories - clear plastic, polystyrene, and blue bag (Table 2). Annotating more examples of object classes with fewer instances will improve the robustness of the model and, in turn, increase the mAP of the model. Augmentation strategies like copy-paste [15] could also be applied to the dataset to improve class imbalance problems. Despite some limitations associated with the dataset, it also offers several advantages. The primary benefit lies in the fact that the images were captured in a realistic setting, replicating the cluttered backgrounds and varying lighting conditions commonly encountered in actual real-world system deployments.

While there are advantages to adopting an autonomous system for detecting contamination at collection, there are some drawbacks to developing such a system. As can be seen in the Fig. 6, in some scenarios, the model detects incorrect objects, multiple objects and misses desired objects due to the highly cluttered background in the dataset. There is an initial expense involved in implementing the computer vision application. As mentioned, the recycling collection truck would need to have a camera attached to it. Furthermore, images would need to be gathered, stored, and annotated, which requires time and associated human annotation costs

need to be taken into account. Utilizing a high-quality camera to capture images with increased detail and clarity could improve the mAP score. The model's performance has been assessed within the confines of commercial and domestic collection routes. Its applicability to other collection routes and environments remains uncertain.

VII. CONCLUSION

In this paper, we have created and utilized the unique Contamination dataset to examine the mAP performance of state-of-the-art algorithms (Faster-RCNN, Cascade-RCNN, Retinanet, and Mask-RCNN with ResNet-101 backbone as well as YOLOv8-x) for object detection and instance segmentation in the context of detecting contamination from recycling waste in commercial waste collection settings. The dataset was generated from video recordings captured during collection routes and was annotated for both object detection and instance segmentation tasks. This work demonstrates the effectiveness of modern computer vision techniques to provide an automated solution to contamination detection at the source. The results showed that employing transfer learning is an effective approach for identifying contamination from recyclable waste in highly cluttered environments, with 0.463 mAP achieved using the YOLOv8-x model with the CSPDarknet backbone. To the best of the authors' knowledge, there are no other viable and autonomous solutions for detecting contamination in real-world settings with dense clutter, as shown in Table 1.

A. NOVELTY

The novelty of the Contamination dataset is underscored by its real-world, in-the-wild characteristics, which distinguishes it from more controlled datasets. It contains data that closely mirrors the operational environments crucial for a computer vision system aimed at detecting contamination in waste materials. The dataset stands out due to the inherent challenges it presents, such as highly overlapping objects, varying object distances, and fluctuating lighting and weather conditions. These real-world variables can alter pixel value distributions, potentially degrading model performance, and collectively create a highly complex challenge for developing effective computer vision methods.

Our work significantly advances the application of computer vision in the waste and recycling sector by addressing the unique challenges of real-world data. Unlike previous studies [43] that primarily utilized controlled datasets, our dataset encompasses a diverse range of environmental conditions, varying object distances, and high object overlap. Similarly, [36] focused on detecting waste materials where the dataset was unbalanced and was tailored to the conditions of recycling plant conveyor belts, limiting the generalizability of their models. By incorporating images from both commercial and residential collection routes, our dataset provides a more comprehensive representation of operational environments. This allows us to train models that are better equipped to handle the complexities of real-world contamination detection. Although direct comparison of mean Average Precision (mAP) values is not feasible due to the differing nature of datasets, our approach demonstrates the potential for improved robustness and adaptability in practical applications, thereby contributing a valuable perspective to the existing body of knowledge in the field.

B. FUTURE WORK

For future work, the dataset can be improved: object classes with fewer instances can be annotated to tackle the problem of class imbalance, and increasing the number of labeled images will improve the performance and robustness of the model. Moreover, higher-resolution images could yield better results in detecting smaller objects. Further data augmentation strategies can be applied to overcome the problem of small and medium-sized objects not being detected.

The automation of contamination detection offers substantial advantages to the waste recycling industry. Firstly, it can enhance workplace safety by minimizing potential employee exposure to hazardous waste by reducing manual handling and inspecting the contamination in recycling waste material. It can also be used to identify individuals contaminating waste at the source, allowing intervention to change customer behavior. This can, in turn, reduce the amount of contaminated recyclable waste that has to be disposed of in landfills rather than recycled. Ultimately, automating contamination detection will lead to cost savings in the overall waste recycling process.

REFERENCES

[1] United States Environmental Protection Agency, Washington, DC, USA. (2018). *National Overview: Facts and Figures on Materials, Wastes and Recycling*. [Online]. Available: https://www.epa.gov/facts-and-figures-about-materials-waste-and-recycling/national-overview-facts-and-figures-materials

[2] Alex. (2019). *The Computer Vision Waste Challenge, 2019*. [Online]. Available: https://www.hr-recycler.eu/the-computer-vision-waste-challenge/

[3] M. A. R. Alif and M. Hussain, "YOLOv1 to YOLOv10: A comprehensive review of YOLO variants and their application in the agricultural domain," 2024, *arXiv:2406.10139*.

[4] D. Bashkirova, M. Abdelfattah, Z. Zhu, J. Akl, F. Alladkani, P. Hu, V. Ablavsky, B. Calli, S. A. Bargal, and K. Saenko, "ZeroWaste dataset: Towards deformable object segmentation in cluttered scenes," 2021, *arXiv:2106.02740*.

[5] L. Biewald. (2020). *Experiment Tracking With Weights and Biases*. [Online]. Available: https://www.wandb.com/

[6] J. Bobulski and M. Kubanek, "Deep learning for plastic waste classification system," *Appl. Comput. Intell. Soft Comput.*, vol. 2021, pp. 1–7, May 2021, doi: 10.1155/2021/6626948.

[7] A. Bochkovskiy, C.-Y. Wang, and H.-Y. Mark Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.

[8] A. L. Brooks, S. Wang, and J. R. Jambeck, "The Chinese import ban and its impact on global plastic waste trade," *Sci. Adv.*, vol. 4, no. 6, Jun. 2018, Art. no. eaat0131.

[9] Z. Cai and N. Vasconcelos, "Cascade R-CNN: Delving into high quality object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6154–6162.

[10] K. Chen et al., "MMDetection: Open MMLab detection toolbox and benchmark," 2019, *arXiv:1906.07155*.

[11] MMYOLO Contributors. (2022). *MMYOLO: OpenMMLab YOLO Series Toolbox and Benchmark*. [Online]. Available: https://github.com/open-mmlab/mmyolo

[12] D. Demetriou, P. Mavromatidis, P. M. Robert, H. Papadopoulos, M. F. Petrou, and D. Nicolaides, "Real-time construction demolition waste detection using state-of-the-art deep learning methods; single-stage vs two-stage detectors," *Waste Manage.*, vol. 167, pp. 194–203, Jul. 2023.
[13] W. Flower. (2013). *What Operation Green Fence Has Meant for Recycling*. Waste360. Accessed: Oct. 18, 2022. [Online]. Available: https://www.waste360.com/business/what-operation-green-fence-has-meant-recycling
[14] M. Fulton, J. Hong, M. J. Islam, and J. Sattar, "Robotic detection of marine litter using deep visual detection models," 2018, *arXiv:1804.01079*.
[15] G. Ghiasi, Y. Cui, A. Srinivas, R. Qian, T.-Y. Lin, E. D. Cubuk, Q. V. Le, and B. Zoph, "Simple copy-paste is a strong data augmentation method for instance segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 2917–2927.
[16] R. Girshick, "Fast R-CNN," 2015, *arXiv:1504.08083*.
[17] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," 2013, *arXiv:1311.2524*.
[18] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," 2014, *arXiv:1406.4729*.
[19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[20] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.
[21] M. Hon and N. M. Khan, "Towards Alzheimer's disease classification through transfer learning," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Nov. 2017, pp. 1166–1169.
[22] J. Hong, M. Fulton, and J. Sattar, "TrashCan: A semantically-segmented dataset towards visual detection of marine debris," 2020, *arXiv:2007.08097*.
[23] Common Objects in Context. (2021). *Detection Evaluation Metrics, 2021*. Accessed: Oct. 18, 2022. [Online]. Available: https://cocodataset.org/detection-eval
[24] G. Jocher, "Ultralytics/YOLOv5: V3.1—Bug fixes and performance improvements," Ultralytics, Frederick, MD, USA, Tech. Rep., Oct. 2020, doi: 10.5281/ZENODO.4154370.
[25] G. Jocher, A. Chaurasia, and J. Qiu. (Jan. 2023). *Ultralytics YOLO*. [Online]. Available: https://github.com/ultralytics/ultralytics
[26] S. Kaza, C. Lisa, P. Bhada-Tata, and F. Van Woerden, *What a Waste 2.0: A Global Snapshot of Solid Waste Management to 2050*. Washington, DC, USA: World Bank, Sep. 2018. [Online]. Available: http://hdl.handle.net/10986/30317
[27] N. V. Kumsetty, A. B. Nekkare, S. S. Kamath, and A. M. Kumar, "TrashBox: Trash detection and classification using quantum transfer learning," in *Proc. 31st Conf. Open Innov. Assoc. (FRUCT)*, Apr. 2022, pp. 125–130.
[28] K. Lamb and A. Morton. (Jul. 2019). *Indonesia Sends Rubbish Back to Australia and Says It's too Contaminated to Recycle*. Guardian. [Online]. Available: https://www.theguardian.com/environment/2019/jul/09/indonesia-sends-rubbish-back-to-australia-and-says-its-too-contaminated-to-recycle
[29] C. Li, L. Li, Y. Geng, H. Jiang, M. Cheng, B. Zhang, Z. Ke, X. Xu, and X. Chu, "YOLOv6 v3.0: A full-scale reloading," 2023, *arXiv:2301.05586*.
[30] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common objects in context," in *Proc. Eur. Conf. Comput. Vis.* Zürich, Switzerland: Springer, 2014, pp. 740–755.
[31] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2999–3007.
[32] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," 2015, *arXiv:1512.02325*.
[33] S. Majchrowska, A. Mikołajczyk, M. Ferlin, Z. Klawikowska, M. A. Plantykow, A. Kwasigroch, and K. Majek, "Deep learning-based waste detection in natural and urban environments," *Waste Manage.*, vol. 138, pp. 274–284, Feb. 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0956053X21006474
[34] W.-L. Mao, W.-C. Chen, H. I. K. Fathurrahman, and Y.-H. Lin, "Deep learning networks for real-time regional domestic waste detection," *J. Cleaner Prod.*, vol. 344, Apr. 2022, Art. no. 131096. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0959652622007284
[35] P. Vingelmann and F. H. P. Fitzek. *CUDA*. NVIDIA CUDA. Accessed: Sep. 8, 2022. [Online]. Available: https://developer.nvidia.com/cuda-11-8-0-download-archive
[36] I. A. Ogrezeanu, C. Suciu, and L. M. Itu, "Automated waste sorting: A comprehensive approach using deep learning for detection and classification," in *Proc. 32nd Medit. Conf. Control Autom. (MED)*, Jun. 2024, pp. 268–273.
[37] H. Panwar, P. K. Gupta, M. K. Siddiqui, R. Morales-Menendez, P. Bhardwaj, S. Sharma, and I. H. Sarker, "AquaVision: Automating the detection of waste in water bodies using deep transfer learning," *Case Stud. Chem. Environ. Eng.*, vol. 2, Sep. 2020, Art. no. 100026.
[38] P. F. Proença and P. Simões, "TACO: Trash annotations in context for litter detection," 2020, *arXiv:2003.06975*.
[39] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," 2016, *arXiv:1612.08242*.
[40] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.
[41] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," 2015, *arXiv:1506.02640*.
[42] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, vol. 28, 2015, pp. 1–14.
[43] P. K. Sarswat, R. S. Singh, and S. V. S. H. Pathapati, "Real time electronic-waste classification algorithms using the computer vision based on convolutional neural network (CNN): Enhanced environmental incentives," *Resour., Conservation Recycling*, vol. 207, Aug. 2024, Art. no. 107651, doi: 10.1016/J.RESCONREC.2024.107651.
[44] D. Soekhoe, P. van der Putten, and A. Plaat, "On the impact of data set size in transfer learning using deep neural networks," in *Proc. 15th Adv. Intell. Data Anal.* Cham, Switzerland: Springer, 2016, pp. 50–60.
[45] A. Wang, H. Chen, L. Liu, K. Chen, Z. Lin, J. Han, and G. Ding, "YOLOv10: Real-time end-to-end object detection," 2024, *arXiv:2405.14458*.
[46] C.-Y. Wang, H.-Y. Mark Liao, I.-H. Yeh, Y.-H. Wu, P.-Y. Chen, and J.-W. Hsieh, "CSPNet: A new backbone that can enhance learning capability of CNN," 2019, *arXiv:1911.11929*.
[47] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," 2022, *arXiv:2207.02696*.
[48] C.-Y. Wang, I.-H. Yeh, and H.-Y. M. Liao, "YOLOv9: Learning what you want to learn using programmable gradient information," 2024, *arXiv:2402.13616*.
[49] S. Wong. (Oct. 2017). *New World Order*. China Scrap Plastics Assoc. [Online]. Available: https://www.recyclingtoday.com/article/national-sword-china-plastics-recycling/
[50] M. Yang and G. Thung, "Classification of trash for recyclability status," Stanford Univ., Stanford, CA, USA, Tech. Rep., 2016.

DISHANT MEWADA received the M.S. degree in artificial intelligence and machine learning from the University of Limerick, Limerick, Ireland, in 2021, where he is currently pursuing the Ph.D. degree in electronic and computer engineering. His research interests include artificial intelligence and computer vision.

CATHAOIR AGNEW received the B.S. degree in financial mathematics and the M.S. degree in artificial intelligence and machine learning from the University of Limerick, Limerick, Ireland, in 2020 and 2021, respectively, where he is currently pursuing the Ph.D. degree in electronic and computer engineering. His research interests include artificial intelligence and computer vision.

EOIN M. GRUA was born in Cork, Ireland, in 1993. He received the B.S. degree in liberal arts and sciences from Amsterdam University College, Amsterdam, The Netherlands, in 2015, the M.S. degree in computer science from Swansea University, Swansea, Wales, in 2016, and the Ph.D. degree in computer science from Vrije Universiteit Amsterdam, Amsterdam, in 2021. In 2021, he was a Research Assistant with the University of Limerick, Limerick, Ireland, where he is currently a Postdoctoral Researcher with the Department of Electronic and Computer Engineering. His research interests include artificial intelligence, software engineering and architecture, and sustainability.

CIARÁN EISING (Senior Member, IEEE) received the B.E. degree in electronic and computer engineering and the Ph.D. degree from NUI Galway, in 2003 and 2010, respectively. From 2009 to 2020, he was a Computer Vision Architect Senior Expert with Valeo. In 2020, he joined the University of Limerick, as an Associate Professor of AI.

PATRICK DENNY (Member, IEEE) received the B.Sc. degree in experimental physics and mathematics from NUI Maynooth, Ireland, in 1993, the M.Sc. degree in mathematics from the University of Galway, Ireland, in 1994, and the Ph.D. degree in physics from the University of Galway (while resident at GFZ Potsdam, Germany), in 2000. From 1999 to 2001, he was a RF Engineer with AVM GmbH, Germany, developing the RF hardware for the first integrated GSM/ISDN/USB modem. After working in supercomputing with Compaq-HP, from 2001 to 2002, he joined Connaught Electronics Ltd., (later Valeo), Galway, Ireland, as the Team Leader of the RF Design. Over the next 20 years, he was the Lead Engineer developing novel RF and imaging systems and led the development of the first mass-production HDR automotive cameras for leading car companies, including Jaguar Land Rover, BMW, and Daimler. In 2010, he became an Adjunct Professor of engineering and informatics with the University of Galway, and a Lecturer of artificial intelligence with the Department of Electronic and Computer Engineering, University of Limerick, Ireland, in 2022. He is currently the Co-Founder and a Committee Member of the IEEE P2020 Automotive Imaging Standards Group, the AutoSens Conference on Automotive Imaging, and the IS&T Electronic Imaging Autonomous Vehicles and Machines (AVM) Conference.

MARK HEFFERNAN received the bachelor's degree in electronic manufacturing from the University of Limerick. He is currently the Software Development Team Lead of AMCS, where he has accumulated over 15 years of experience working with the IoT, vehicle technology, and enterprise software solutions. He is particularly interested in applying machine learning techniques to enhance efficiency in the waste management industry, aiming to drive significant environmental improvements through innovative technological applications.

KEN TIERNEY is currently the Product Manager of AMCS Group. He is involved in AMCS IoT technology products, where he is applying automation combined with computer vision and AI to drive sustainability in the waste and recycling industry. He has been with AMCS for over 14 years and his background is in electronics and sustainability energy management.

PEPIJN VAN DE VEN received the M.Sc. degree in electronic engineering from Eindhoven University of Technology, The Netherlands, in 2000, and the Ph.D. degree in artificial intelligence for autonomous underwater vehicles from the University of Limerick, in 2005. He is currently a Professor of artificial intelligence with the University of Limerick.

ANTHONY SCANLAN received the B.Sc. degree in experimental physics from the National University of Ireland, Galway, Ireland, in 1998, and the M.Eng. and Ph.D. degrees in electronic engineering from the University of Limerick, Limerick, Ireland, in 2001 and 2005, respectively. He is currently a Senior Research Fellow with the Department of Electronic and Computer Engineering, University of Limerick, and has been the Principal Investigator of several research projects in the areas of signal processing and data converter design. His current research interests include artificial intelligence, computer vision, and industrial and environmental applications.