Waste Management 178 (2024) 35–45

Contents lists available at ScienceDirect

Waste Management

journal homepage: www.elsevier.com/locate/wasman

Research Paper

CODD: A benchmark dataset for the automated sorting of construction and demolition waste

Demetris Demetriou [a,*], Pavlos Mavromatidis [b,c], Michael F. Petrou [a], Demetris Nicolaides [b,c]

[a] Department of Civil & Environmental Engineering, University of Cyprus, Nicosia 1303, Cyprus
[b] Frederick Research Centre, Nicosia 1036, Cyprus
[c] Frederick University, Nicosia 1036, Cyprus

ARTICLE INFO

Keywords:
Construction and demolition waste
Object detection
Benchmark dataset
Automated sorting
Bounding box detection
Instance segmentation

ABSTRACT

This study presents the Construction and Demolition Waste Object Detection Dataset (CODD), a benchmark dataset specifically curated for the training of object detection models and the full-scale implementation of automated sorting of Construction and Demolition Waste (CDW). The CODD encompasses a comprehensive range of CDW scenarios, capturing a diverse array of debris and waste materials frequently encountered in real-world construction and demolition sites. A noteworthy feature of the presented study is the ongoing collaborative nature of the dataset, which invites contributions from the scientific community, ensuring its perpetual improvement and adaptability to emerging research and practical requirements. Building upon the benchmark dataset, an advanced object detection model based on the latest bounding box and instance segmentation YOLOV8 architecture is developed to establish a baseline performance for future comparisons. The CODD benchmark dataset, along with the baseline model, provides a reliable reference for comprehensive comparisons and objective assessments of future models, contributing to progressive advancements and collaborative research in the field.

1. Introduction

The efficient management of Construction and Demolition Waste (CDW) poses a significant and multifaceted challenge for the construction industry, given its profound environmental and economic impacts. CDW represents over one-third of all waste in the European Union, highlighting its status as the largest waste stream by volume and emphasizing the urgency of addressing its impacts (Bilsen et al., 2018). Recognizing the crucial role of CDW in environmental sustainability, the EU (Directive 2008/98/EC) has placed a strong emphasis on the recovery, reuse, and recycling of CDW within its Circular Economy agenda. This stance is critical in steering the construction sector towards more sustainable practices, especially considering the favourable comparison of recycling over landfilling and incineration in terms of global warming potential (Ortiz et al., 2010).

Despite legislative and environmental initiatives, recycling CDW remains challenging, mainly due to concerns over recycled material purity. These concerns, often stemming from the variability in waste composition and the presence of contaminants, lead to reduced market demand for recycled products (Al-Raqeb et al., 2023; Medina et al., 2015; Vegas et al., 2015). Adding to these complexities is the practice of off-site sorting, which is often preferred by contractors despite some countries mandating source separation of CDW (Menegaki and Damigos, 2018; Ulubeyli et al., 2017). In light of these challenges, refining the CDW sorting process becomes crucial, as it is instrumental in guaranteeing the quality of the recycled materials.

CDW sorting spans from basic centralized and decentralized systems (Bao et al., 2020) to more sophisticated automated robotic solutions. Traditional sorting systems rely on heavy machinery, including feed hoppers, crushing systems, magnetic separators, screens and conveyors to separate CDW (Hu et al., 2019). These systems are often complemented at the final stages of the process with manual sorting to further refine the material purity (Demetriou et al., 2023; Huang et al., 2002; Hyvarinen et al., 2020). Manual sorting of CDW, nonetheless, is often deemed to be inconsistent, costly, unreliable, and hazardous to the workers involved in the process (Davis et al., 2021; Sarc et al., 2019). To overcome these drawbacks, the integration of robotic systems in CDW sorting is gaining traction, driven by recent advancements in machine/deep learning and convolutional neural networks. This shift towards automated (and autonomous) sorting is transforming waste

* Corresponding author.
E-mail address: demetriou.c.demetris@ucy.ac.cy (D. Demetriou).

https://doi.org/10.1016/j.wasman.2024.02.017
Received 24 September 2023; Received in revised form 16 January 2024; Accepted 12 February 2024
Available online 19 February 2024

management practices, aiming to replace human involvement with more accurate, efficient robotic systems in both on-site and off-site sorting operations (Bosoc et al., 2021; Chen et al., 2022; Lukka et al., 2014; Sarc et al., 2019; Wang et al., 2020; Wang et al., 2019; Xiao et al., 2020). This transition is seen as a pivotal step in enhancing waste management efficiency and contributing to sustainable practices in the construction sector. The efficacy of these robotic systems, however, hinges crucially on their ability to accurately locate and classify different types of waste, a challenge being actively addressed by recent advances in machine learning and artificial intelligence. This area of research is rapidly evolving, with significant implications for the optimization and efficiency of automated sorting systems.

Early automation attempts using robots focused on traditional machine learning methods to enhance waste detection accuracy and effectiveness (Bonifazi et al., 2019; Wang et al., 2019; Xiao et al., 2019). While pioneering, traditional machine learning often faced limitations in efficiency, accuracy, and generalizability, especially in complex waste sorting scenarios (Rad et al., 2017; Yu et al., 2020; Zhang et al., 2021; Zhao et al., 2022) as they required explicit knowledge of the features to be identified, making them less adaptable to the varied and unpredictable nature of CDW materials.

As the limitations of traditional approaches became evident, the field gradually shifted towards adopting more advanced techniques. This evolution is marked by recent contributions such as those by (Chen et al., 2022; Demetriou et al., 2023; Dong et al., 2022; Lin et al., 2023; Lux et al., 2023), which addressed specific challenges in the domain of CDW detection and sorting. For example, Demetriou et al. (2023) focused on the complexities of stacking and adherence of CDW samples by employing and comparing different single-stage and two-stage CNN detection architectures. Lux et al. (2023) proposed the RACNET architecture for the precise classification and accurate estimation of mass, class, and binary masks from 2D images of recycled aggregates. Chen et al. (2022) developed an instance segmentation model that considers aspects such as object size and pose to enhance the grasping efficiency of robots. Dong et al. (2022) introduced the Boundary-Aware Transformer (BAT) framework for fine-grained composition recognition of construction waste mixtures. Various other studies, including the studies of (Li et al., 2022; Lin et al., 2022; Song et al., 2022; Zhou et al., 2022), significantly contributed to improving the detection accuracy, expanded the recognition of diverse CDW materials, and made important strides toward the real-world implementation of automated CDW sorting systems. The reader is referred to Table A1 in the appendix for a comprehensive overview of recent studies related to CDW detection and sorting for further reference.

At the same time, it is imperative to acknowledge a notable challenge that persists - the scarcity of comprehensive and standardized benchmark datasets for training and evaluating CDW detection models. The absence of such datasets undermines the ability to perform objective comparisons and assessments of different models, thereby hindering the overall advancement of the field. Acknowledging this pivotal gap, this study aims to develop a much-needed benchmark in order to establish a solid foundation that paves the way for future research and innovation in automated CDW sorting. A benchmark dataset can serve as an invaluable tool, facilitating rigorous evaluations and objective comparisons of various models, methodologies, and techniques proposed in the literature.

Indeed, benchmark datasets have proven to be highly beneficial in various domains. In domestic waste classification, the TrashNet dataset proposed by Yang and Thung (2016) has advanced research in this area as it serves as a baseline for the development of various object detection models on CNN architectures with great reported success (Aral et al., 2018; Mao et al., 2021). Extended datasets like the ImageNet (Deng et al., 2009) have played a crucial role in driving advancements in computer vision. In natural language processing, the Stanford Sentiment Treebank (SST) dataset (Socher et al., 2013) has been influential for sentiment analysis tasks. The KITTI dataset (Geiger et al., 2013) has been instrumental in autonomous driving research, providing data for tasks such as object detection and scene understanding. In healthcare, the MIMIC-III dataset (Johnson et al., 2016) has enabled the development and evaluation of machine learning models using electronic health records. These benchmark datasets provide standardized data and evaluation metrics, allowing researchers to compare and benchmark their models against common references, thus driving progress and fostering innovation in their respective fields.

In the domain of CDW sorting, however, existing studies utilize datasets of varying sizes, compositions, and acquisition methodologies (the reader is referred to Table A1 in the appendix). Each study relies on its own dataset, limiting the direct comparison of model performance. This lack of standardization impedes collaboration and collective progress within the field. Against this backdrop, the development of an extended, expansive and openly accessible benchmark dataset, as envisaged in this study, represents an invaluable and transformative resource for researchers. The proposed dataset establishes a standardized reference, providing a common ground to evaluate the performance of various CDW detection models and benchmark their proposed methods. As with other scientific domains, this dataset will facilitate the reproducibility of results and will encourage the sharing of ideas, techniques, and advancements in CDW detection methodologies, encouraging collective scientific progress.

The main objective of this study is to establish the foundation for future advancements in automated and autonomous CDW sorting technologies. This is achieved by addressing the notable gap in comprehensive and standardised data resources through two main avenues: the curation of a benchmark dataset, namely the Construction and Demolition Waste Object Detection Dataset (CODD), and the development of an advanced object detection model to serve as a baseline for future comparisons. The establishment of a standardized reference for evaluating CDW detection models will encourage collaboration, reproducibility, and collective advancement in the domain. Through this work, the authors envisage creating a dynamic and adaptable resource by continuously updating the CODD benchmark with contributions from the scientific community so that it aligns with emerging research needs and practical requirements.

2. Research contributions

This work aspires to set the stage for the widespread adoption of automated and autonomous robotic CDW sorting systems by making contributions in several key areas:

1) Benchmark dataset creation: This study sets to bridge the existing data gap by developing a benchmark dataset, CODD, for training and testing bounding box and instance segmentation object detection models in automated CDW sorting systems.

2) Baseline detection model: This study introduces a state-of-the-art detection model to serve as a standardized reference for the evaluation of future CDW sorting models and methods, providing a baseline for technological advancements in the field.

3) Collaborative research: This study opens up new avenues for collaborative research and innovation in the field of CDW sorting by actively inviting researchers to participate, contribute/exchange data, and assess their respective models, methodologies, and techniques.

3. Methodology

3.1. Dataset acquisition

To create an accurate and comprehensive benchmark dataset, this study adopted a systematic process to acquire and annotate all necessary data. Drawing inspiration from the European Commission's waste classification for CDW (European Waste Catalogue, 2001), a wide range of CDW categories were collected from a recycling facility in Cyprus (Fig. 1) to develop the training, validation and testing datasets in the

Fig. 1. (a) CDW recycling facility and (b) piles of manually sorted CDW from where the samples were extracted.

CODD. The datasets comprised ten distinct categories, namely bricks, concrete, tiles, wood, pipes, plastics, general waste (including nylon and plastic bottles commonly found at construction and demolition sites), foaming insulation, stones, and plasterboards. In alignment with Europe's waste classification, which includes categories such as "concrete, bricks, tiles and ceramics," "wood, glass and plastic," and "metals," this study aimed to capture the CDW types most commonly encountered in construction sites. Notably, certain categories such as metals, glass and asbestos-containing construction materials were omitted from the dataset. This decision was based on the ease of removing metals from the waste stream using magnets and the special handling requirements associated with glass and asbestos-containing materials. Images containing samples of items belonging to the ten categories of CDW considered in the development of the CODD are presented in Fig. A1 in the appendix.

During the data collection process, emphasis was placed on inclusivity rather than imposing specific size criteria on the selected samples. To this end, a wide range of CDW object sizes were selected to accommodate the diverse waste sorting methods employed by any potential end-user of the dataset. To achieve this, an all-inclusive approach was adopted, whereby all CDW objects that could fit on the conveyor belt were recorded. This decision ensured the dataset's versatility and accounted for scenarios where larger-sized items might be handled using specialised grippers or other sorting mechanisms. To illustrate the range of sizes included in the dataset, Fig. A2 in the appendix displays the mean pixel area per object class. It is worth noting that these values were calculated using the surveyor's formula (Braden, 1986) applied to the ground truth polygons (the reader is referred to section 2.2 for details regarding ground truth development) of each object.

Additionally, attention was placed on recording samples that exhibited stacking and adherence, as this is known to present challenges in CDW object detection. The study of Demetriou et al. (2023) showcased the degradation of object detection performance on stacked and adhered samples, highlighting the importance of incorporating such variations in the dataset to develop robust CDW detection models. Finally, to preserve the fidelity of the dataset, all collected samples were utilized in their original state and as received from the recycling facility. This approach ensured that the inherent characteristics and properties of the CDW objects encountered in real-world waste streams were accurately represented.

The dataset acquisition process involved capturing images using a HIKROBOT MV-CA023-10GC camera overlooking a conveyor belt on which the CDW objects were placed. A total of 3,129 images containing a total of 16,545 CDW samples were recorded. Each image possessed an original size of $1920 \times 1200 \times 3$, with the colour channels being represented in RGB. The data collection phase encompassed different lighting conditions, including both artificial and ambient lighting setups, emulating the real-world variability observed during CDW sorting operations. For a detailed insight into the distribution of the samples across the dataset, Fig. A3 in the appendix illustrates the number of samples per image.

3.2. Dataset annotation

The acquired images were annotated with great care to ensure the dataset's accuracy. This approach involved two independent expert annotators and a single supervisor who reviewed and refined the annotations, correcting any misclassifications and enhancing the precision of the bounding boxes. The initial step of this process, wherein each object was labelled with bounding boxes, was conducted using MATLAB's image labeller software, as illustrated in Fig. A4 in the appendix.

Secondly, to facilitate and expedite the labelling procedure, teacher models were used to predict approximate classifications and bounding box locations. The reader is referred to Demetriou et al. (2023) and references therein for more details regarding the deployment of teacher models. In images featuring densely stacked and adhered CDW samples (as depicted in Fig. 2), extra care was placed to guarantee the precision of bounding box annotation by both the human annotators and the teacher models. This was achieved through enhanced supervision to ensure precise classification and bounding box location in these complex scenarios.

Thirdly, the samples were further annotated with segmentation masks using a semi-automated method. Accordingly, bounding box annotations were converted to instance segmentation masks with the use of the very recent Segment Anything Model (SAM) (Kirillov et al., 2023). SAM uses a combination of convolutional neural networks and advanced image processing algorithms to analyse the visual features and spatial relationships within the images, enabling precise object segmentation. The code files for the implementation of SAM can be accessed through: https://github.com/facebookresearch/segment-anything. Fig. 2 presents an example of annotated objects using bounding boxes and instance segmentation masks generated by the SAM, illustrating simultaneously the progression from single-object, single-class images to multi-object, multi-class, heavily stacked and adhered images.

It is noted that each image in the dataset is accompanied by a single file in the standardized Visual Objects Classes (VOC) XML format. Accordingly, each XML file contains detailed annotations, including the object class, bounding box coordinates (x_{min}, y_{min}, x_{max}, and y_{max})

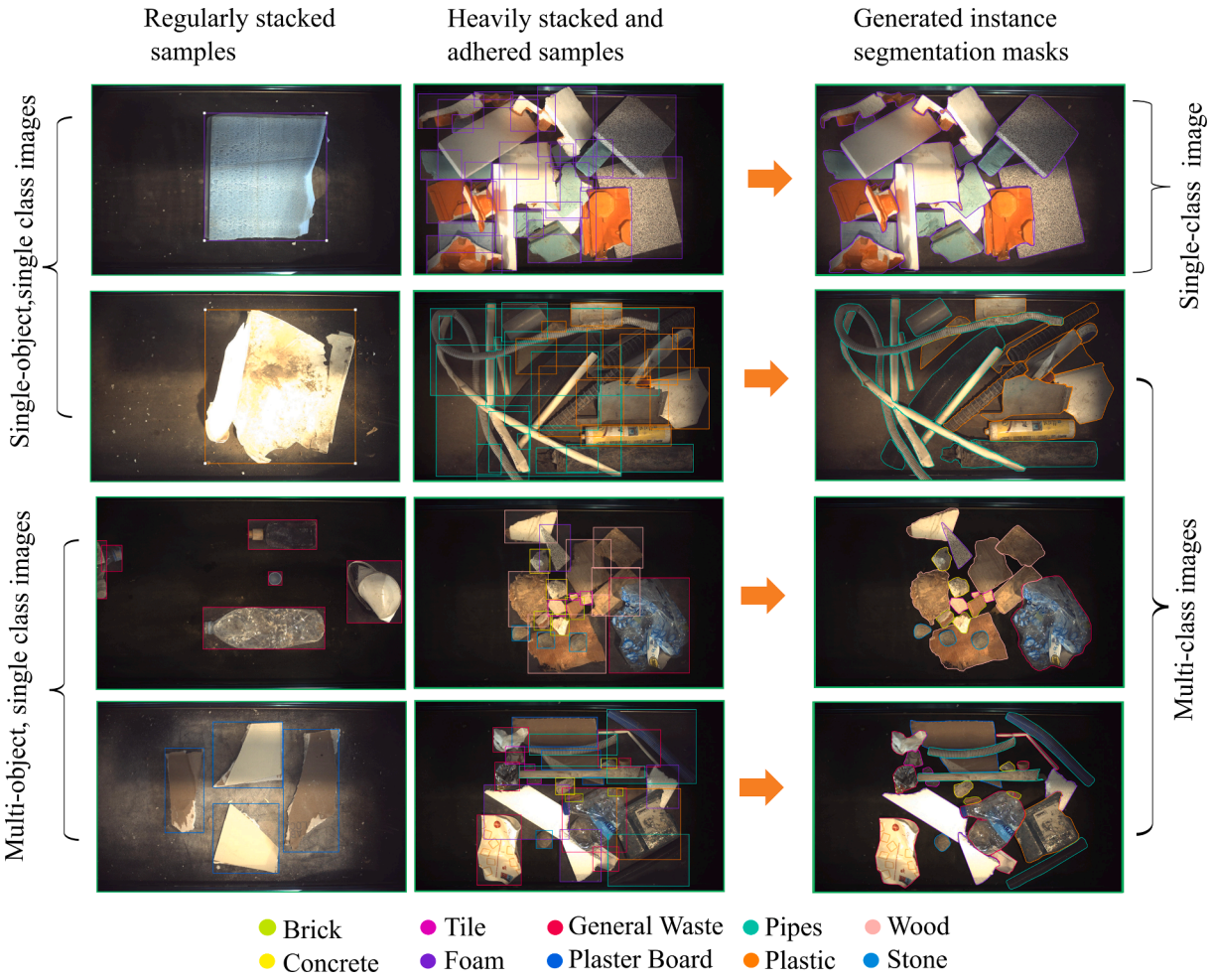

Fig. 2. An illustrative example of annotated objects in the dataset showcasing the utilization of both bounding boxes (left) and SAM generated instance segmentation masks (right).

defining the enclosing rectangle, as well as the coordinates of the polygons (x_1, y_1, x_2, y_3, …, x_n, y_n) outlining the precise boundaries of the objects. To aid visualization, a website application has been developed and can be accessed at https://coddannotator.streamlit.app/. This application enables users of the CODD to visualize the annotated images interactively. Instructions on how to use the application are available at the provided link.

3.3. Dataset split

The images in the benchmark dataset were split into training, validation, and testing subsets with an approximate 70-15-15 % ratio to

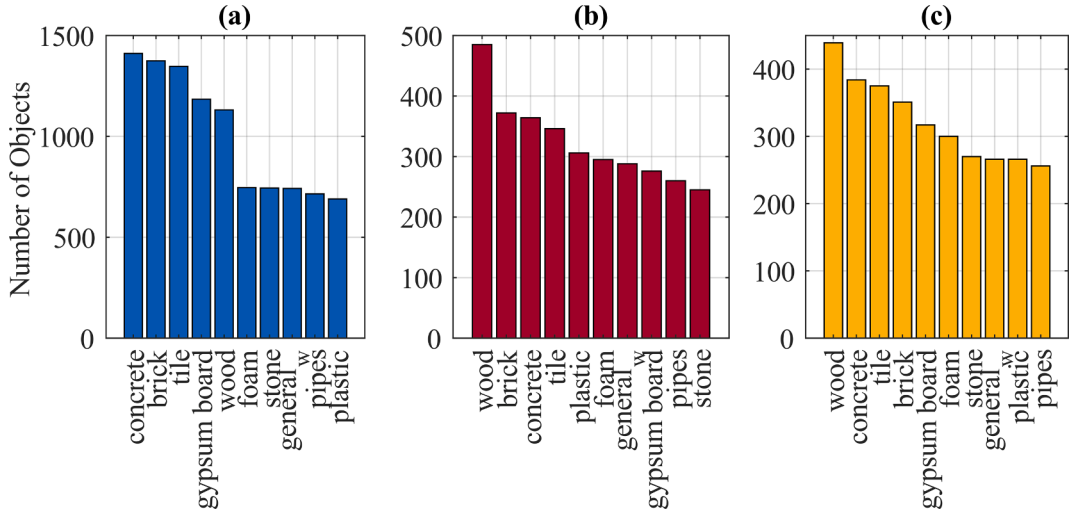

Fig. 3. Distribution of samples per object class in the (a) training, (b) validation, and (c) testing set of the benchmark dataset.

allow for a sufficient amount of training data while reserving separate subsets for parameter tuning and evaluation. The final distribution of samples per class within the benchmark dataset is graphically depicted in Fig. 3. To address the inherent variability found in certain CDW object categories, such as bricks, concrete, tiles, and wood, the dataset was specifically designed to include a larger number of unique samples for these materials. This deliberate increase in the number of unique samples ensured a comprehensive coverage of the diverse characteristics and features inherent to these CDW types. Increasing the number of samples for these specific materials, the dataset captures their diverse and complex characteristics, facilitating more comprehensive training of CDW detection models.

All training, validation, and testing subsets (images and annotations) are located in dedicated folders within the benchmark repository (https://data.mendeley.com/datasets/wds85kt64j/3), providing easy access to researchers and users.

Finally, a critical aspect of the CODD dataset that is worth highlighting is the provision of a fixed testing set for all users of the benchmark so that future evaluations of methods and models can be performed on an equal basis. At the same time, it is worth noting that while the testing set should always remain fixed, there exists a degree of flexibility regarding the utilization of the validation set provided in the benchmark repository. This discretionary aspect allows users of the CODD dataset to exercise their judgment in determining whether to incorporate the validation set into their specific research framework since there might be instances where a validation set may not be deemed necessary or may be superseded by alternative validation strategies.

3.4. Dataset perpetuity and expansion

Ensuring the longevity and availability of the benchmark dataset is key in facilitating ongoing research and promoting the reproducibility of results. This study recognizes the significance of maintaining the dataset's perpetuity and adopts a proactive approach to address this aspect. Firstly, to ensure the benchmark's long-term accessibility and preservation, the dataset has been deposited in a dedicated repository (https://data.mendeley.com/datasets/wds85kt64j/3). This repository serves as a centralized and secure location that allows researchers and users to easily access, download, and utilize the dataset for their CDW detection experiments. Secondly, to ensure the expandability of the dataset, the authors commit to providing updates and improvements based on community feedback, contributions and emerging research needs.

Indeed, it is essential to acknowledge that the CODD benchmark, while comprehensive, represents merely a fractional glimpse into the vast pool of potential images that a multitude of researchers have the capacity to amass, with a varying arsenal of cameras, resolutions, distances from objects, perspective changes and an array of other contributing factors that can further enrich the diversity in the dataset. Certainly, with active engagement with the scientific community and a commitment to welcoming contributions, the CODD benchmark will continue to evolve, accommodating new CDW categories, equipment variations, and real-world environmental conditions, bringing the domain one step closer to realizing full-scale implementation.

The authors kindly request contributors to reach out for specific details regarding the submission of their contributions. Upon contact, contributors will receive comprehensive guidelines outlining key aspects to consider before submitting, including annotation format and metadata documentation.

3.5. Training and fine-tuning the baseline models

To initiate the training process, the YOLOV8, a state-of-the-art single-stage and anchor-free object detection architecture known for its efficiency and accuracy (Guo et al., 2023; Hussain, 2023; Kim et al., 2023; Terven et al., 2023) was selected. For a deeper understanding of the contextual relevance of the YOLO architecture to the particular task, readers are encouraged to refer to the study of Demetriou et al. (2023), which presents a comprehensive comparison of various single-stage and two-stage detection architectures specifically in the context of CDW detection.

In order to establish a performance baseline, a range of bounding box detection models were developed on the YOLOV8 architecture at different scales, namely YOLOV8n, YOLOV8s, YOLOV8m, YOLOV8l, and YOLOV8x. These models were characterized by varying parameters, with totals of 3.2, 11.2, 25.9, 43.7, and 68.2 million parameters, respectively. The study extended to the training of instance segmentation models, denoted as YOLOV8n-seg, YOLOV8s-seg, YOLOV8m-seg, YOLOV8l-seg, and YOLOV8x-seg, featuring 3.4, 11.8, 27.3, 46.0, and 71.8 million parameters. It is noted that YOLOV8 segmentation models repurpose the YOLOV8 object detection architecture, i.e., the model also predicts both bounding box coordinates and class probabilities, even though the primary focus lies on generating segmentation masks.

All models underwent training on Google's Collaboratory platform using an Nvidia Tesla T4 GPU with 16 GB of memory. Throughout the training phase, all models underwent iterative optimization of their parameters. A synergistic combination of techniques was employed to enhance the model's performance; Stochastic gradient descent (SGD) optimization was used to minimize the model's loss function by adjusting weights based on gradients computed from mini-batches of training samples. During the training process, batch normalization was used as a means of preventing overfitting by stabilizing the learning process through normalization of the output of each layer within the network, reducing the likelihood of overly complex and specialized representations. Pretrained weights for all YOLOV8 models were utilized as a starting point to expedite the training and convergence process. These weights can be accessed from the official YOLO GitHub repository (Jocher et al., 2023). For convenience and further reference, these weights are also available in the CODD's repository (https://data.mendeley.com/datasets/wds85kt64j/3).

In pursuit of robust and versatile model performance, two distinct data augmentation strategies were implemented. Firstly, an offline data augmentation strategy was employed as a preprocessing step to generate augmented training samples. This augmentation pipeline aimed to diversify the training dataset, enhancing the model's ability to generalize to various real-world scenarios. To this end, each image in the training set underwent a series of augmentations, resulting in the generation of three distinct augmented images per original input image, with a total of 5958 images (1928 x 3) being generated for training. This augmentation process encompassed several key transformations. A cropping technique was applied with a range of zoom levels, spanning from a minimal zoom of 0 % to a maximum zoom of 35 %. This augmentation intends to mimic different perspectives and distances of the CDW materials on the conveyor belt. Moreover, variations in brightness were introduced, encompassing fluctuations within a range of -10 % to $+10$ %. This augmentation enables the model to robustly handle varying levels of illumination. Mosaic augmentation was also incorporated into the pipeline. This technique involves the synthesis of multiple images to create a cohesive mosaic that simulates complex and cluttered scenes. Finally, bounding box shearing was applied to introduce spatial deformations. This augmentation introduced horizontal and vertical shear transformations within the range of $\pm 15^{\circ}$, effectively simulating the distortion that objects might undergo due to perspective changes. It is important to emphasize that the authors arrived at these augmentation procedures and associated values through experimentation. Finally, Fig. 4 presents a collection of images generated as a result of the above augmentations.

Secondly, online data augmentation was applied using the YOLOV8 bag-of-freebies at each training iteration. This augmentation pipeline introduced variations into CDW objects, effectively increasing the model's ability to handle diverse CDW scenarios and environmental factors. Accordingly, the YOLO data augmentation procedure consisted of various techniques, including HSV-Hue augmentation, HSV-

Fig. 4. Collection of images resulting from the offline data augmentation procedure applied during preprocessing.

Saturation augmentation, image HSV-Value augmentation, image translation, scaling, random rotation at 90 degrees, and mosaic augmentation. These augmentations were implemented following the recommended values provided in the YOLOV8 GitHub repositories (Jocher et al., 2023). It is worth noting that all models were trained on two input image resolutions, specifically 320 × 320 and 640 × 640, to evaluate scale robustness. A selection of the key training parameters is illustrated in Table 1 below.

Table 1
Key training parameters for the YOLOV8 models.

Parameter	Value
Epochs	25
Batch size	16
Input image size	320 × 320 and 640 × 640 pixels
Learning rate	0.01
Learning schedule	Linear
Momentum	0.937
Weight decay	0.001
Warmup epochs	3
Warmup momentum	0.8
Intersection over union	0.7
Augmentations at each iteration	HSV_Hue = 0.015, HSV_Saturation = 0.7, HSV_Value = 0.4, translate = 0.1, scale = 0.5, fliplr = 0.5, mosaic = true

4. Evaluation metrics

4.1. Accuracy metrics

The evaluation of object detection models, particularly within the context of a benchmark dataset, requires the use of standardized metrics and algorithm-independent metrics to facilitate meaningful comparisons and promote scientific rigor. In this spirit, the CODD benchmark employs the well-established evaluation metrics of Average Precision (AP) and mean Average Precision (mAP) to assess the performance of the detection models and enable objective comparisons with future research. These metrics, as exemplified by major benchmarks such as the PASCAL VOC (Everingham et al., 2009) and MS COCO (Lin et al., 2014), are chosen for their universal applicability across various object detection methodologies. Average Precision (AP) serves as a fundamental metric in object detection evaluation and measures the detection precision at various recall levels. It is calculated independently for each considered object class, allowing for class-specific performance analysis. The mean Average Precision (mAP), on the other hand, provides a comprehensive evaluation of the detector's overall performance by averaging the AP values across all object classes, providing a single value that reflects the model's effectiveness in detecting CDW classes.

The calculation of AP involves the generation of a precision-recall curve, which captures the trade-off between precision and recall. Precision is defined as the ratio of true positive detections to the sum of true

positive and false positive detections (eq.1).

$$p = \frac{TP}{TP + FP} \tag{1}$$

Recall represents the ratio of true positive detections to the total number of ground truth objects (eq.2). By varying the intersection-over-union (IOU) threshold, which determines the level of overlap required for a detection to be considered correct, the detector's performance at different levels of localization accuracy is assessed.

$$r = \frac{TP}{TP + FN} \tag{2}$$

To calculate AP, integration of the precision over the recall range is performed, as denoted in eq.3.

$$AP = \int_0^1 p(r)dr \tag{3}$$

The precision and recall values are obtained by analysing the classification predictions and the IOU between the predicted bounding boxes (B_p) and ground truth bounding boxes (B_{gt}) for each image in the testing dataset (eq.4).

$$IOU = \frac{area(B_p \cap B_{gt})}{area(B_p \cup B_{gt})} \tag{4}$$

True positive (TP) detections are those with an IOU greater than the specified threshold and correct classification. In contrast, false positive (FP) detections are incorrect predictions with an IOU lower than the threshold. False negative (FN) refers to ground truth objects that were not detected by the model.

In the evaluation of CDW detection models, the use of specific IOU thresholds is essential for consistent and standardized comparisons. In line with established evaluation practices in object detection, the study calculates AP at several IOU thresholds, including the baseline (Pascal VOC metric) threshold of 0.50. Additionally, the study considers multiple IOU thresholds ranging from 0.50 to 0.95 at intervals of 0.05 to provide a comprehensive assessment of the detector's performance across a wide range of localization accuracies. The AP values at these thresholds are averaged to obtain $AP_{50:95}$ for each object class. Furthermore, mAP_{50} and $mAP_{50:95}$ represent the averaged AP values at different thresholds, providing a holistic evaluation of the detector's performance across all considered object classes.

The use of AP and mAP metrics in evaluating CDW detection models within the framework of a benchmark dataset and baseline model ensures objective and standardized comparisons. The authors would like to highlight that adhering to these metrics can facilitate future research, enabling researchers to benchmark their own CDW detection methods and models against the presented baseline model.

4.2. Inference speed metrics

The evaluation of inference speed is conducted by measuring the time taken to process a single image throughout three distinct stages: pre-processing, inference, and post-processing. The total inference speed (pre-processing, inference and post-processing) of the baseline model is measured on a Tesla T4 16 GB GPU.

5. Results and discussion

5.1. Model performance evaluation

In this section, the evaluation of two types of models investigated in this study, namely bounding box and instance segmentation models, is presented. The performance of these models is rigorously assessed on the metrics of Average Precision (AP) and mean Average Precision (mAP) in order to establish a performance baseline that will enable objective comparisons and benchmarking against future research. Each model was trained on the training set and evaluated on the dedicated validation and testing sets. The $mAP_{50:95}$, representing the averaged AP values across different Intersection over Union (IOU) thresholds (ranging from 0.50 to 0.95 at intervals of 0.05), is used as the overall performance indicator. Below, the $mAP_{50:95}$ achieved by the bounding box and instance segmentation models as a function of the inference speed across different model scales and input image resolutions is graphically presented in Fig. 5 and Fig. 6 (complemented by Table A2 and Table A3 in the appendix). For the completeness of the investigation and benchmarking purposes, analytical and per object class detection results for all the models developed herein are provided as supplementary material to this paper and can be accessed through the benchmark repository (https://data.mendeley.com/datasets/wds85kt64j/3). Finally, PyTorch (.pt) files of all the models trained on CODD have been made available, ensuring accessibility for further research and application in the field.

As evident from the results depicted in Fig. 5 and Fig. 6, some general remarks can be made. Firstly, it is apparent that models trained on a higher image resolution naturally achieve superior performance over their counterparts. Secondly, there is an evident positive correlation between model scale and detection accuracy, with the larger model scale (YOLOV8x) outperforming their smaller counterparts both with respect to bounding box (Fig. 5) and instance segmentation detection (Fig. 6), at the expense of inference speed. While this observation might seem trivial, this increase in performance as a function of model scale across the whole spectrum is a primary indicator of the absence of overfitting. Indeed, this observation suggests that within the context of the CODD benchmark, it is apparent that the problem domain is sufficiently complex to require models of increased complexity. This observation in turn suggests the suitability of the CODD as a benchmark dataset on which very complex model architectures can be implemented.

On the contrary, if overfitting was present, one would expect a point at which increasing the model's complexity (scale) would lead to diminishing returns, potentially even a decrease in performance on the validation and/or test data. The absence of degradation in performance between the validation and testing sets (as evident both in Fig. 5 and Fig. 6), in addition to the absence of performance degradation as a function of the model scale, reinforces the argument for good model generalisation and against overfitting, suggesting that the models developed herein exhibit the robustness necessary for a potential expansion in real-world applications.

Furthermore, the comparison between YOLOV8n and YOLOV8s, as well as the YOLOV8n-seg and YOLOV8s-seg, particularly at lower (320 × 320) image resolutions yields that small increases in model complexity (e.g., increasing the model scale from YOLOV8n to YOLOV8s) can result to substantial gains in mAP, as evident from the steep gradient between these two points. Conversely, larger model scales beyond the YOLOV8s and YOLOV8s-seg exhibit a diminishing rate of increase in performance.

Fig. 7 illustrates the per-class results of the bounding box and instance segmentation models achieving the best overall detection performance, namely YOLOV8x and YOLOV8x-seg. It is noteworthy, and as anticipated, that YOLOV8x demonstrates superior performance compared to the YOLOV8x-seg model across all classes. This can be attributed to the additional complexity associated with the latter task (e.g., instance segmentation). Additionally, a commendable overall ($mAP_{50:95}$) performance is attained for most classes, signifying the model's efficacy in accurately localizing and classifying objects. However, it is also noted that the object classes 'plastic' and 'pipes' achieve slightly lower mAP scores, suggesting potential areas for further optimization, data enrichment or data augmentation to enhance performance in these specific categories.

For visualisation, Fig. 8 presents a collection of CDW objects detected using both bounding boxes and instance segmentation masks using the

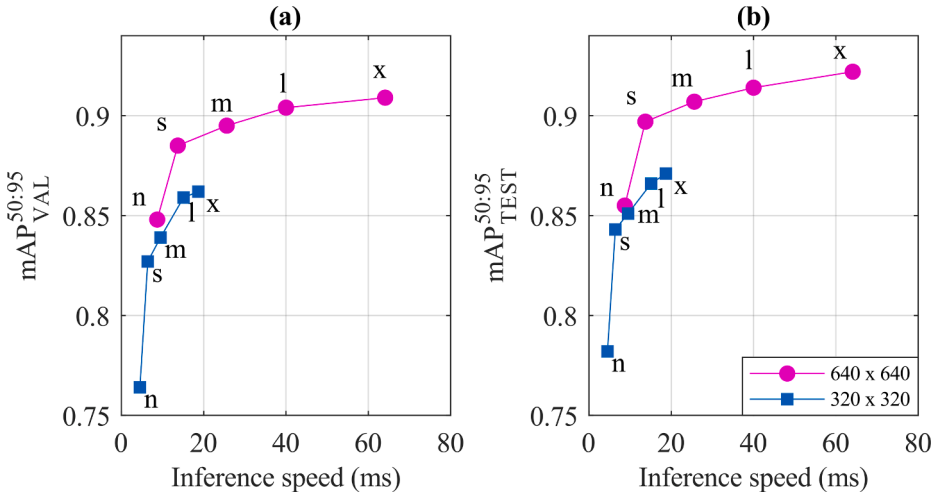

Fig. 5. Performance comparison of different scales of the YOLOV8 bounding box models on the (a) validation and (b) testing set at an input image resolution 320 × 320 and 640 × 640.

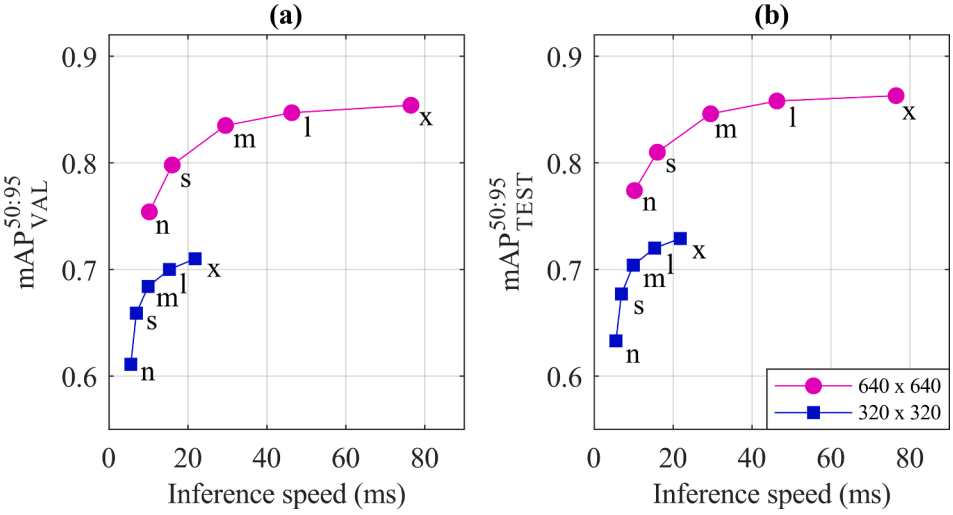

Fig. 6. Performance comparison of different scales of the YOLOV8 instance segmentation models on the (a) validation and (b) testing set at an input image resolution of 320 × 320 and 640 × 640 (subscripts '-seg' e.g., YOLOV8n-seg, YOLOV8s-seg etc. have been immitted for figure clarity).

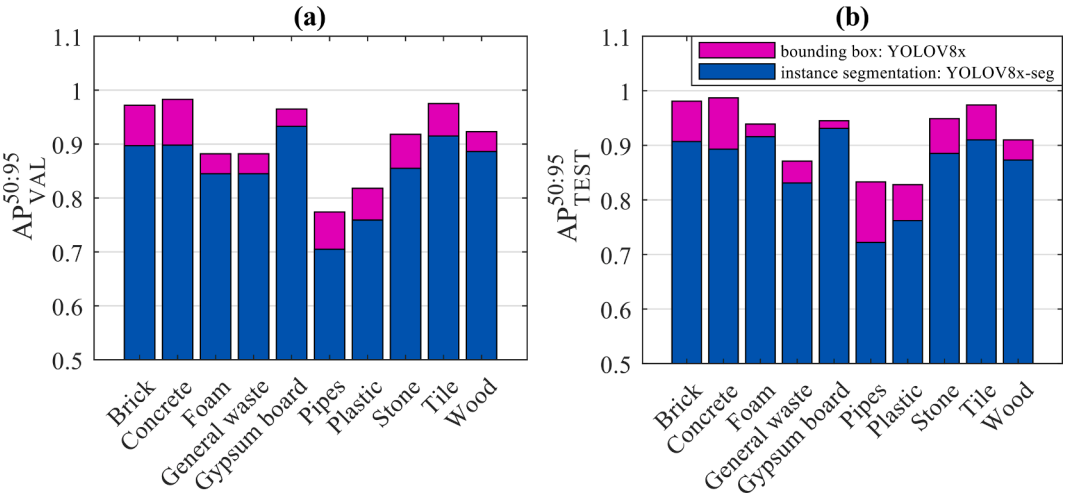

Fig. 7. Bounding box and instance segmentation performance of the YOLOV8X model per object class on (a) validation and (b) testing set.

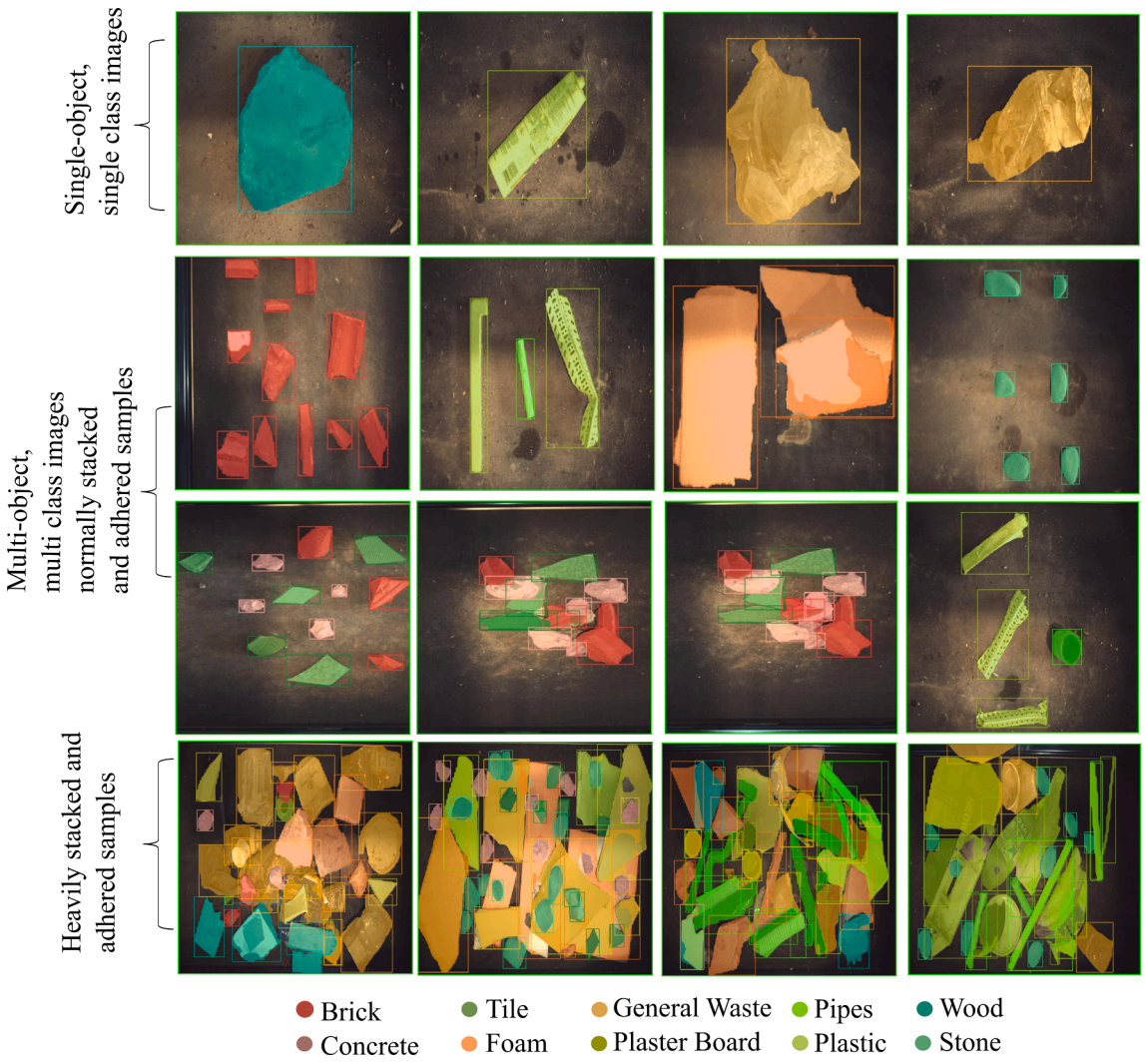

Fig. 8. YOLOV8x-seg (640 × 640) bounding boxes and instance segmentation masks detected on selected testing set images.

YOLOV8x-seg model. It is observed that most objects in the figure are accurately identified and contoured, demonstrating the capability of the YOLOV8x-seg model to identify CDW accurately. For a qualitative assessment of the various scale instance segmentation models trained on the CODD dataset, the authors developed an interactive tool accessible at https://yolov8inferencetool.streamlit.app/. This web application enables users to interactively explore inference results from different model scales, offering valuable insights into their performance. Specifically, the tool can be used to demonstrate the efficacy of progressively larger model scales in handling complex scenarios, such as samples exhibiting stacking and adherence, within the CODD testing dataset.

6. Conclusions

This study addresses a notable gap, the absence of a dedicated benchmark dataset in the field of Construction and Demolition Waste (CDW) sorting, by introducing a comprehensive benchmark dataset specifically tailored to the training and evaluation of object detection (bounding box and instance segmentation) models. The Construction and Demolition Waste Object Detection Dataset (CODD), developed in this study, captures an assortment of CDW materials commonly encountered in real-world construction and demolition sites. The dataset consists of 10 CDW categories, including bricks, concrete, tiles, wood, pipes, plastics, general waste, foaming insulation, stones, and plaster boards, ensuring an accurate and realistic representation of the diverse waste materials encountered during CDW sorting operations.

For the development of the CODD, a diverse array of RGB images were recorded and annotated in the standardized Visual Objects Classes (VOC) XML format, ensuring user-friendly access for future usage. The provided files contain both bounding box coordinates and instance segmentation masks generated using the very recent Segment Anything Model (SAM), which facilitated the efficient conversion of bounding box coordinates to polygon coordinates. Both images and annotations are structured and available in dedicated folders within the benchmark repository (https://data.mendeley.com/datasets/wds85kt64j/3).

The study introduced an advanced object detection model based on the YOLOV8 architecture, establishing a performance baseline for future comparisons and assessments. In an attempt to extract maximum performance from all the developed models, a meticulous data augmentation procedure was introduced as a preprocessing step to enhance the diversity and complexity of the images in the training set. A systematic evaluation of the developed models across various scales and image resolutions was performed and the findings exhibited a clear correlation between model scale, image resolution, and detection accuracy. Most importantly, a crucial consistency in performance between the validation and testing sets was observed, suggesting the absence of overfitting as well as the robustness and generalisability of all developed models. As anticipated, the largest model scale, YOLOV8x and YOLOV8x-seg, showcased the highest detection accuracy, naturally at the expense of inference speed. For disclosure, transparency and future referencing,

analytical model results and all the YOLOV8 models (in PyTorch.pt format) developed herein are provided as supplementary material to this paper.

The authors acknowledge the dynamic nature of research in this domain and eagerly anticipate contributions from the scientific community. With an open invitation for contributions to the scientific community, the authors are confident that the CODD is poised to grow, encompassing an even broader spectrum of CDW categories, equipment variations, and real-world environmental conditions. The authors strongly believe that a collective effort is required to get one step closer to the realization of widespread and effective CDW sorting implementations.

Funding Source

This research was funded by the Cyprus Government through the Research & Innovation Foundation (RIF) and the European Regional Development Fund (ERDF) in the frame of the RESTART Programmes for Research, Technological Development and Innovation 2016–2020 (Project Budget: €1,098,880, Project Contract Number: INTEGRATED/0918/0052).

CRediT authorship contribution statement

Demetris Demetriou: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Writing – original draft, Writing – review & editing. **Pavlos Mavromatidis:** Data curation, Investigation, Supervision, Validation, Visualization. **Michael F. Petrou:** Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing – review & editing. **Demetris Nicolaides:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to express their sincere gratitude to the Cyprus Research & Innovation Foundation (RIF) and the European Regional Development Fund (ERDF), for funding the research project entitled “Development of an Innovative Insulation Fire Resistant Façade from the Construction and Demolition Wastes” (Contract Number: INTEGRATED/0918/0052). Additionally, the authors gratefully acknowledge Frederick University and the University of Cyprus for providing access to their facilities and data. Finally, the authors would like to express their gratitude to Resource Recovery Cyprus (RRC) Ltd, who provided the recycled material to support the efforts in this study.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.wasman.2024.02.017.

References

Al-Raqeb, H., Ghaffar, S.H., Al-Kheetan, M.J., Chougan, M., 2023. Understanding the challenges of construction demolition waste management towards circular construction: Kuwait Stakeholder’s perspective. Cleaner Waste Syst. 4, 100075.

Aral, R.A., Keskin, Ş.R., Kaya, M., Haciömeroğlu, M., 2018. Classification of trashnet dataset based on deep learning models, 2018 IEEE International Conference on Big Data (Big Data). IEEE, pp. 2058-2062.

Bao, Z., Lee, W.M.W., Lu, W., 2020. Implementing on-site construction waste recycling in Hong Kong: Barriers and facilitators. Sci. Total Environ. 747, 141091.

Bilsen, V., Kretz, D., Padilla, P., Van Acoleyen, M., Van Ostaeyen, J., Izdebska, O., Hansen, M., Bergmans, J., Szuppinger, P.J.E.C.B., Belgium, 2018. Development and implementation of initiatives fostering investment and innovation in construction and demolition waste recycling infrastructure. 206.

Bonifazi, G., Capobianco, G., Serranti, S., 2019. Hyperspectral Imaging and Hierarchical PLS-DA Applied to Asbestos Recognition in Construction and Demolition Waste. Applied Sciences 9.

Bosoc, S., Suciu, G., Scheianu, A., Petre, I., 2021. Real-time sorting system for the Construction and Demolition Waste materials, 2021 13th International Conference on Electronics, Computers and Artificial Intelligence (ECAI). IEEE, pp. 1-6.

Braden, B., 1986. The Surveyor’s area formula. Coll. Math. J. 17, 326–337.

Chen, X., Huang, H., Liu, Y., Li, J., Liu, M., 2022. Robot for automatic waste sorting on construction sites. Autom. Constr. 141.

Davis, P., Aziz, F., Newaz, M.T., Sher, W., Simon, L., 2021. The classification of construction waste material using a deep convolutional neural network. Automation in Construction 122.

Demetriou, D., Mavromatidis, P., Robert, P.M., Papadopoulos, H., Petrou, M.F., Nicolaides, D., 2023. Real-time construction demolition waste detection using state-of-the-art deep learning methods; single-stage vs two-stage detectors. Waste Manage. 167, 194–203.

Deng, J., Dong, W., Socher, R., Li, L.J., Kai, L., Li, F.-F., 2009. ImageNet: A large-scale hierarchical image database, 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248-255.

Dong, Z., Chen, J., Lu, W., 2022. Computer vision to recognize construction waste compositions: a novel boundary-aware transformer (BAT) model. J. Environ. Manage. 305, 114405.

European Waste Catalogue, 2001. Commission Decision 2001/118/EC of 16 January 2001 amending Decision 2000/532/EC as regards the list of wastes, p. 47.

Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A., 2009. The pascal visual object classes (VOC) challenge. Int. J. Comput. Vis. 88, 303–338.

Geiger, A., Lenz, P., Stiller, C., Urtasun, R., 2013. Vision meets robotics: The KITTI dataset. Int. J. Robot. Res. 32, 1231–1237.

Guo, J., Lou, H., Chen, H., Liu, H., Gu, J., Bi, L., Duan, X.J.S.r., 2023. A new detection algorithm for alien intrusion on highway. 13, 10667.

Hu, K., Chen, Y., Naz, F., Zeng, C., Cao, S., 2019. Separation studies of concrete and brick from construction and demolition waste. Waste Manage. 85, 396–404.

Huang, W.-L., Lin, D.-H., Chang, N.-B., Lin, K.-S., 2002. Recycling of construction and demolition waste via a mechanical sorting process. Resour. Conserv. Recycl. 37, 23–37.

Hussain, M., 2023. YOLO-v1 to YOLO-v8, the Rise of YOLO and Its Complementary Nature toward Digital Manufacturing and Industrial Defect Detection, Machines.

Hyvarinen, M., Ronkanen, M., Karki, T., 2020. Sorting efficiency in mechanical sorting of construction and demolition waste. Waste Manage. Res. 38, 812–816.

Jocher, G., Chaurasia, A., Qiu, J., 2023. YOLO by Ultralytics.

Johnson, A.E.W., Pollard, T.J., Shen, L., Lehman, L.-w.H., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Anthony Celi, L., Mark, R.G., 2016. MIMIC-III, a freely accessible critical care database. Scientific Data 3, 160035.

Kim, J.H., Kim, N., Won, C.S., 2023. High-Speed Drone Detection Based On Yolo-V8, ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1-2.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.-Y.J.a.p.a., 2023. Segment anything.

Li, J., Fang, H., Fan, L., Yang, J., Ji, T., Chen, Q., 2022. RGB-D fusion models for construction and demolition waste detection. Waste Manag. 139, 96–104.

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

Lin, K., Zhou, T., Gao, X., Li, Z., Duan, H., Wu, H., Lu, G., Zhao, Y., 2022. Deep convolutional neural networks for construction and demolition waste classification: VGGNet structures, cyclical learning rate, and knowledge transfer. J. Environ. Manage. 318, 115501.

Lin, K., Zhao, Y., Zhou, T., Gao, X., Zhang, C., Huang, B., Shi, Q., 2023. Applying machine learning to fine classify construction and demolition waste based on deep residual network and knowledge transfer. Environ. Dev. Sustain. 25, 8819–8836.

Lukka, T.J., Tossavainen, T., Kujala, J.V., Raiko, T., 2014. Zenrobotics recycler–robotic sorting using machine learning, Proceedings of the International Conference on Sensor-Based Sorting (SBS). Citeseer, p. 1.

Lux, J., Lau Hiu Hoong, J.D., Mahieux, P.-Y., Turcry, P., 2023. Classification and estimation of the mass composition of recycled aggregates by deep neural networks. Computers in Industry 148.

Mao, W.-L., Chen, W.-C., Wang, C.-T., Lin, Y.-H., 2021. Recycling waste classification using optimized convolutional neural network. Resources, Conservation and Recycling 164.

Medina, C., Zhu, W., Howind, T., Frías, M., De Sánchez Rojas, M.I., 2015. Effect of the constituents (asphalt, clay materials, floating particles and fines) of construction and demolition waste on the properties of recycled concretes. Constr. Build. Mater. 79, 22–33.

Menegaki, M., Damigos, D., 2018. A review on current situation and challenges of construction and demolition waste management. Curr. Opin. Green Sustainable Chem. 13, 8–15.

Ortiz, O., Pasqualino, J.C., Castells, F., 2010. Environmental performance of construction waste: comparing three scenarios from a case study in Catalonia, Spain. Waste Manage. 30, 646–654.

Rad, M.S., von Kaenel, A., Droux, A., Tieche, F., Ouerhani, N., Ekenel, H.K., Thiran, J.-P., 2017. A Computer Vision System to Localize and Classify Wastes on the Streets. In:

Liu, M., Chen, H., Vincze, M. (Eds.), Computer Vision Systems. Springer International Publishing, Cham, pp. 195–204.
Sarc, R., Curtis, A., Kandlbauer, L., Khodier, K., Lorber, K.E., Pomberger, R.J.W.M., 2019. Digitalisation and intelligent robotics in value chain of circular economy oriented waste management – a review. 95, 476-492.
Socher, R., Perelygin, A., Wu, J.Y., Chuang, J., Manning, C.D., Ng, A.Y., Potts, C., 2013. Recursive deep models for semantic compositionality over a sentiment treebank. EMNLP 1631, 1631–1642.
Song, L., Zhao, H., Ma, Z., Song, Q., 2022. A new method of construction waste classification based on two-level fusion. PLoS One 17, e0279472.
Terven, J., Cordova-Esparza, D.J.a.p.a., 2023. A comprehensive review of YOLO: From YOLOv1 to YOLOv8 and beyond.
Ulubeyli, S., Kazaz, A., Arslan, V.J.P.E., 2017. Construction and demolition waste recycling plants revisited: management issues. 172, 1190-1197.
Vegas, I., Broos, K., Nielsen, P., Lambertz, O., Lisbona, A.J.C., Materials, B., 2015. Upgrading the quality of mixed recycled aggregates from construction and demolition waste by using near-infrared sorting technology. 75, 121-128.
Wang, Z., Li, H., Yang, X.J.J.o.B.E., 2020. Vision-based robotic system for on-site construction and demolition waste sorting and recycling. 32, 101769.
Wang, Z., Peng, B., Huang, Y., Sun, G., 2019. Classification for plastic bottles recycling based on image recognition. Waste Manage. 88, 170–181.
Xiao, W., Yang, J., Fang, H., Zhuang, J., Ku, Y., 2019. Development of online classification system for construction waste based on industrial camera and hyperspectral camera. PLoS One 14, e0208706.
Xiao, W., Yang, J., Fang, H., Zhuang, J., Ku, Y., Zhang, X., 2020. Development of an automatic sorting robot for construction and demolition waste. Clean Techn. Environ. Pol. 22, 1829–1841.
Yang, M., Thung, G., 2016. Classification of Trash for Recyclability Status; CS229 Project Report. Stanford University, Stanford, CA, USA.
Yu, B., Wang, J., Li, J., Lu, W., Li, C.Z., Xu, X., 2020. Quantifying the potential of recycling demolition waste generated from urban renewal: A case study in Shenzhen, China. J. Clean. Prod. 247.
Zhang, Q., Yang, Q., Zhang, X., Bao, Q., Su, J., Liu, X., 2021. Waste image classification based on transfer learning and convolutional neural network. Waste Manag. 135, 150–157.
Zhao, X., Yang, Y., Duan, F., Zhang, M., Jiang, G., Yan, X., Cao, S., Zhao, W., 2022. Identification of construction and demolition waste based on change detection and deep learning. Int. J. Remote Sens. 43, 2012–2028.
Zhou, Q., Liu, H., Qiu, Y., Zheng, W., 2022. Object Detection for Construction Waste Based on an Improved YOLOv5 Model. Sustainability 15.