Waste Management 190 (2024) 149–160

Contents lists available at ScienceDirect

Waste Management

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

Optimizing waste handling with interactive AI: Prompt-guided segmentation of construction and demolition waste using computer vision

Diani Sirimewan [a], Nilakshan Kunananthaseelan [b], Sudharshan Raman [c], Reyes Garcia [d], Mehrdad Arashpour [e,*]

[a] Department of Civil Engineering, Faculty of Engineering, Monash University, Melbourne, Australia
[b] Department of Electrical and Computer Systems Engineering, Faculty of Engineering, Monash University, Melbourne, Australia
[c] Civil Engineering Discipline, School of Engineering, Monash University, Malaysia
[d] School of Engineering, University of Warwick, Coventry, United Kingdom
[e] Department of Civil Engineering, Monash University, Melbourne, Australia

ARTICLE INFO

Keywords:
Construction and demolition waste
Automated waste recognition
Prompt-guided segmentation
Waste monitoring and sorting
Computer vision
Automation

ABSTRACT

Optimized and automated methods for handling construction and demolition waste (CDW) are crucial for improving the resource recovery process in waste management. Automated waste recognition is a critical step in this process, and it relies on robust image segmentation techniques. Prompt-guided segmentation methods provide promising results for specific user needs in image recognition. However, the current state-of-the-art segmentation methods trained for generic images perform unsatisfactorily on CDW recognition tasks, indicating a domain gap. To address this gap, a user-guided segmentation pipeline is developed in this study that leverages prompts such as bounding boxes, points, and text to segment CDW in cluttered environments. The adopted approach achieves a class-wise performance of around 70 % in several waste categories, surpassing the state-of-the-art algorithms by 9 % on average. This method allows users to create accurate segmentations by drawing a bounding box, clicking, or providing a text prompt, minimizing the time spent on detailed annotations. Integrating this human–machine system as a user-friendly interface into material recovery facilities enhances the monitoring and processing of waste, leading to better resource recovery outcomes in waste management.

1. Introduction

The management of increasing amounts of construction and demolition waste (CDW) and the scarcity of natural resources present significant challenges for the construction industry (Demetriou et al., 2024; Laadila et al., 2021; Véliz et al., 2022). More than 10 billion tons of CDW are produced annually, most ending up in landfills (Yazdani et al., 2021). However, landfill waste disposal poses significant threats to sustainable development and human health, causing irreversible ecological damage, increased land degradation risks, water pollution, and climate change (Li and Zhang, 2024; Yong et al., 2023). As a solution, the CDW can be processed into recycled materials, which can be reused in new construction applications without being disposed of in landfills. This involves a series of processes, such as transporting waste to the material recovery facilities (MRFs), quality inspection, sorting, reprocessing, secondary market operations, and residue disposal (Tennakoon et al., 2022). Recycling is a widely used method of reprocessing waste materials like concrete, aggregates, metal, timber and plastic for manufacturing new products (Demetriou et al., 2023). For instance, studies have demonstrated the potential of using recycled aggregates to construct road bases and replace virgin aggregates in concrete mixes (Di Maria et al., 2016; Jayasinghe et al., 2019; Li et al., 2022). Thus, CDW can be transformed into recycled materials, reducing waste sent to landfills and decreasing the demand for virgin materials. This gives CDW a second life, diverting it from landfills and extracting value from materials that would have otherwise been discarded. This repurposing reduces environmental pollution and promotes sustainability (Islam et al., 2019).

MRFs play a critical role in the resource recovery process (Tennakoon et al., 2022). They are a vital link in the supply chain, connecting demolition contractors who transport CDW from sites to end users in the secondary market (Chileshe et al., 2019). They manually inspect the composition of incoming waste and sort the mixed waste for

* Corresponding author at: Monash University, Department of Civil Engineering, VIC 3800, Australia.
E-mail addresses: diani.sirimewan@monash.edu (D. Sirimewan), Nilakshan.Kunananthaseelan@monash.edu (N. Kunananthaseelan), sudharshan.raman@monash.edu (S. Raman), Reyes.Garcia@warwick.ac.uk (R. Garcia), mehrdad.arashpour@monash.edu (M. Arashpour).

https://doi.org/10.1016/j.wasman.2024.09.018
Received 31 July 2024; Received in revised form 13 September 2024; Accepted 20 September 2024
Available online 24 September 2024
0956-053X/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Nomenclature

CDW	construction and demolition waste
CLIP	contrastive language-image pre-training
CNN	convolutional neural network
CRD	construction, renovation and demolition
CV	computer vision
DSC	Dice-Sørensen coefficient
GPU	graphics processing unit
IoU	intersection over union
MAE	masked autoencoder
MLP	multilayer perceptron
MRFs	materials recovery facilities
RCNN	region-based convolutional neural network
RGB-D	red-green-blue-depth
SAM	segment anything model
SEEM	segment everything everywhere all at once model
SOTA	state-of-the-art
SSD	single-shot detector
ViT	vision transformer
YOLACT	you only look at coefficients
YOLO	you only look once

recycling. The labor cost is identified as a primary driver of the processing cost of incoming waste at MRFs. This issue has been further exacerbated by the high labor rates (Sirimewan et al., 2024a). The multi-stage monitoring and sorting systems at MRFs require workers to inspect the incoming waste and segregate them at various stages in the process. Fig. 1 illustrates the overview of waste handling at MRFs.

During the first stage of the monitoring system, all incoming waste loads are inspected on the weighbridge at the point of entry to MRFs. Trained weighbridge operators monitor each truckload of waste that enters the MRFs to ascertain the suitability of the materials for processing. The waste loads suitable for processing are forwarded to the tipping area. As the second stage of the monitoring system, specialized teams of trained employees examine the waste loads as they are unloaded at the tipping area (Tennakoon, 2024). The conventional approach for categorizing CDW primarily depends on manual sorting based on visual inspection. Subsequently, machinery clears the remaining waste from the sorting area, directing them to the designated stockpiles (Sirimewan et al., 2024b). However, this manual procedure is laborious, costly, and less efficient, posing hazards to workers (Demetriou et al., 2024; Li et al., 2024). Thus, it is necessary to devise intelligent methods to automate waste recognition and subsequent handling.

The use of automated processes to facilitate labor-intensive tasks has proven to be not only feasible but also highly beneficial in waste management (Lu et al., 2022). Many MRFs have fully automated crushing plants and semi-automated sorting processes, such as using magnets to remove metals and air-blowers to remove lightweight materials like plastics. These methods lead to safer and more efficient waste handling (Lu and Chen, 2022; Tennakoon, 2024). The automated systems can be computer-controlled, utilizing cameras, computers, and robotic arms to monitor, sort, and process waste materials. The utilization of AI for automated waste handling is gaining popularity with the progressions in computer vision-based approaches (Arashpour, 2023; Moral et al., 2022).

Computer vision (CV) advancements have been remarkable in the field of waste management (Jin et al., 2023; Li and Zhang, 2024). These developments utilize deep learning and image processing techniques to address the challenges of waste identification, classification, and segregation (Dodampegama et al., 2024; Pitakaso et al., 2024). CV technologies have introduced automated systems capable of real-time waste detection and segmentation. Waste detection methods have evolved from basic object recognition algorithms to more advanced techniques involving convolutional neural networks (CNNs) and their variants (Majchrowska et al., 2022; Nežerka et al., 2024). Segmentation methods, which involve partitioning an image into meaningful segments, have seen remarkable improvements using deep learning architectures such as DeepLabv3+ (Chen et al., 2018), U-Net (Ronneberger et al., 2015), Mask R-CNN (He et al., 2017), and fully convolutional networks. These models can delineate waste object boundaries within an image, enabling precise extraction and classification of various waste types (Lu et al., 2022). Additionally, Transformer-based models, such as vision transformers (ViTs) (Dosovitskiy et al., 2020), have demonstrated superior performance on various benchmarks, often surpassing state-of-the-art (SOTA) CNN architectures. Their scalability and flexibility make them well-suited for large-scale datasets that accommodate diverse and heterogeneous nature (Dong et al., 2022). Table 1 below illustrates recent research on automating CDW recognition and sorting using CV methods.

Despite technological advancements, using SOTA image segmentation methods for waste recognition is still limited. Creating accurate segmentations based solely on CV without human intervention is challenging due to the heterogeneous nature of CDW streams (Fu et al., 2024). On the one hand, automation of the waste handling process is complex, mainly because the current technological capabilities

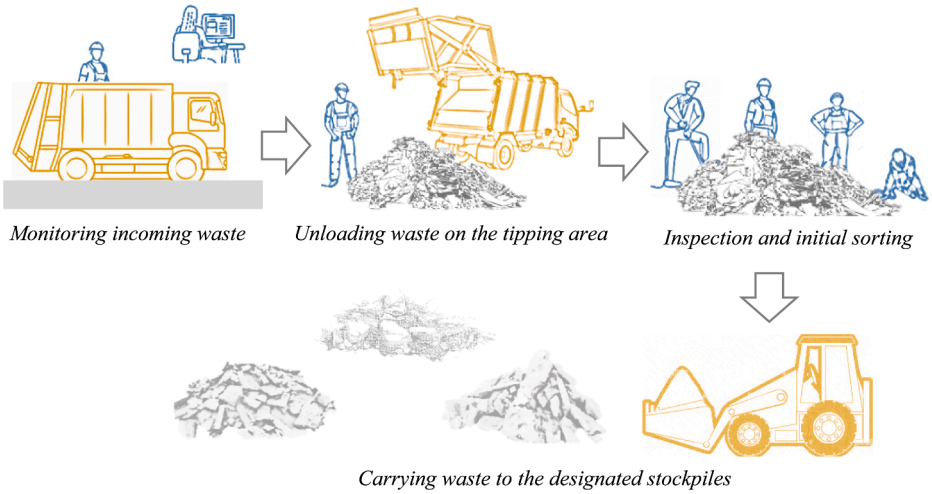


Fig. 1. Waste handling process at MRFs.

Table 1
Recent research on automation of CDW recognition and sorting using CV methods.

Authors/ Source	Research Purpose	Computer vision methods
Demetriou et al. (2024)	CDW object detection for automated sorting	Instance segmentation of waste objects using YOLOV8
Wu et al. (2024)	Particle size distribution monitoring in CDW recycling	Instance segmentation of 3D-laser-triangulation images using U-Net-based multi-star algorithm
Li et al. (2024)	Building decoration waste sorting system using robots	Detection of waste objects using YOLOX
Prasad and Arashpour (2024)	Segmentation of recyclables from cluttered CDW streams using RGB-D_L fusion model	Instance segmentation of waste objects using Mask-RCNN-based (late concatenation) fusion model
Sirimewan et al. (2024a)	Recognising the composition of mixed CRD waste in-the-wild	Semantic segmentation of waste using DeepLabv3+ and U-Net
Sirimewan et al. (2024b)	Semi-supervised segmentation of CDW in-the-wild	Semantic segmentation of waste using adversarial dual view networks
Yong et al. (2023)	Automatic identification of CDW landfills in large-scale areas	Semantic segmentation of remote sensing imagery using DeepLabv3+
Chen et al. (2023)	Augmented reality-enabled human-robot collaboration for construction waste sorting	Morphology-based object segmentation to locate waste objects in real-time images
Wang et al. (2023)	Automated segmentation of recycling materials for semantic understanding in construction	Segmentation of images to recognize recyclables using Transformer-based architectures
Demetriou et al. (2023)	Real-time localisation and classification of CDW for robotic sorting	Detection of waste objects using single-stage (SSD, YOLO) and two-stage detectors (Faster-RCNN)
Li et al. (2022)	Real-time detection of CDW with RGB-D fusion models	Instant segmentation of waste objects using Mask-RCNN-based fusion models
Lu et al. (2022)	Recognising the composition of cluttered construction waste mixtures	Semantic segmentation of waste materials using DeepLabv3+
Dong et al. (2022)	Fine-grained recognition of cluttered construction waste composition	Semantic segmentation of waste materials with boundary refinement using Transformers
Na et al. (2022)	Automated waste segmentation and classification system for recycling CDW	Instant segmentation using YOLACT
Chen et al. (2022)	Automatic waste sorting at construction sites using robots	Instance segmentation of waste using Mask R-CNN

constrain the automation of heavy equipment operations, such as excavators and front-end loaders (Tennakoon, 2024). On the other hand, the lack of publicly available datasets significantly hinders the application of CV models in real-world CDW management (Sirimewan et al., 2024b). Creating such datasets is time-consuming and requires detailed annotation and domain-specific expertise, limiting these models' scalability and practical deployment. Therefore, integrating an effective human–machine interface can be proposed to manage resources at MRFs and for efficient and precise data annotations.

Recent advancements in image segmentation have introduced foundational models like Segment Anything Model (SAM) (Kirillov et al., 2023) and Segment Everything Everywhere All at Once (SEEM) (Zou et al., 2024), which demonstrate exceptional versatility and performance across diverse segmentation tasks (Ma et al., 2024). They can perform segmentation tasks based on user-guided prompts, making them highly adaptable across diverse segmentation tasks (Wei et al., 2024). They operate on a prompt-guided mechanism where users can provide different types of inputs, such as points, boxes, scribbles, or text prompts, to guide the region of interest. This interactive approach allows flexible and accurate segmentation by adapting to the tasks not seen during training (Wu and Xu, 2024). However, these models are primarily trained on large-scale generic image datasets, which presents limitations when applied to specific downstream applications (Ma et al., 2024; Mazurowski et al., 2023). CDW involves diverse materials in cluttered environments, where different materials are mixed in various forms, shapes, and sizes (Demetriou et al., 2024; Lu and Chen, 2022). The existing models trained on generic datasets do not capture the diversity of waste materials, leading to poor generalization and suboptimal results when applied to CDW tasks. The study fills this gap by adapting the SOTA models to the domain of CDW. A segmentation pipeline is developed named 'PromSeg-Waste', which is a flexible, prompt-based segmentation method that leverages user inputs such as bounding boxes, points, and text to identify and segment waste materials in real-time. The method can be integrated into MRFs as a human–machine interface, streamlining waste monitoring, quality inspection and sorting processes. This improves the efficiency of waste handling by optimizing cost, time, and loabor through the combined strengths of AI and human expertise. Additionally, PromSeg-Waste allows users to significantly reduce time spent on detailed annotations by generating precise segmentations through simple actions, such as drawing a bounding box, clicking, or providing a text prompt.

2. Materials and methods

2.1. Overview of the PromSeg-Waste model

The SOTA segmentation foundation models use a sophisticated image encoder-prompt encoder-mask decoder architecture. The image encoder processes input images and computes an image embedding. The prompt encoder embeds the prompts into a form the model can understand. Subsequently, the information from these two sources is combined using a mask decoder, which enables segmented masks to be made as the output (Kirillov et al., 2023). Despite being extensively trained on large-scale datasets of generic images, they struggle when there is a substantial domain shift in downstream tasks (Cheng et al., 2023; Xiong et al., 2024). The literature has overlooked promptable segmentation in CDW and its applications. Hence, we developed PromSeg-Waste by adapting the SOTA segmentation foundation models as a pipeline for prompt-guided CDW segmentation. Fig. 2(a) illustrates the architecture of PromSeg-Waste.

The PromSeg-Waste model utilizes Transformer architecture, as in the SOTA models known for success in natural language processing and image processing tasks. It integrates a masked autoencoder (MAE) pre-trained Vision Transformer (ViT)-based image encoder, which extracts key features from images that process user inputs like bounding boxes, points, or text, and a mask decoder that generates segmentation maps. The base ViT model includes 12 Transformer layers, each comprising a multi-head self-attention block and a multilayer perceptron (MLP) block with layer normalization (norm) as shown in Fig. 2(b). They are responsible for serving the model focus on different parts of the image. The image encoder divides the input image into patches and flattens them into vectors (embeddings). The Transformer then processes these vectors to understand relationships between different patches. The pre-trained MAE ensures the embeddings retain essential image information even when parts are obscured.

The prompt encoder takes user inputs and converts them into vector embeddings that represent the position of the input in the image. Sparse prompts (or simple inputs), such as bounding boxes and points, are encoded based on their spatial positions within the image. The dense prompts (or detailed inputs), such as masks, are downscaled through convolutional layers to match the image. The model then combines the image and input information. Afterwards, the mask decoder creates an accurate segmentation map using a dual cross-attention mechanism, which efficiently merges the user's prompt with image data. Finally, the model produces the output by processing these elements through additional layers, ensuring the segmentation matches the user's input. This

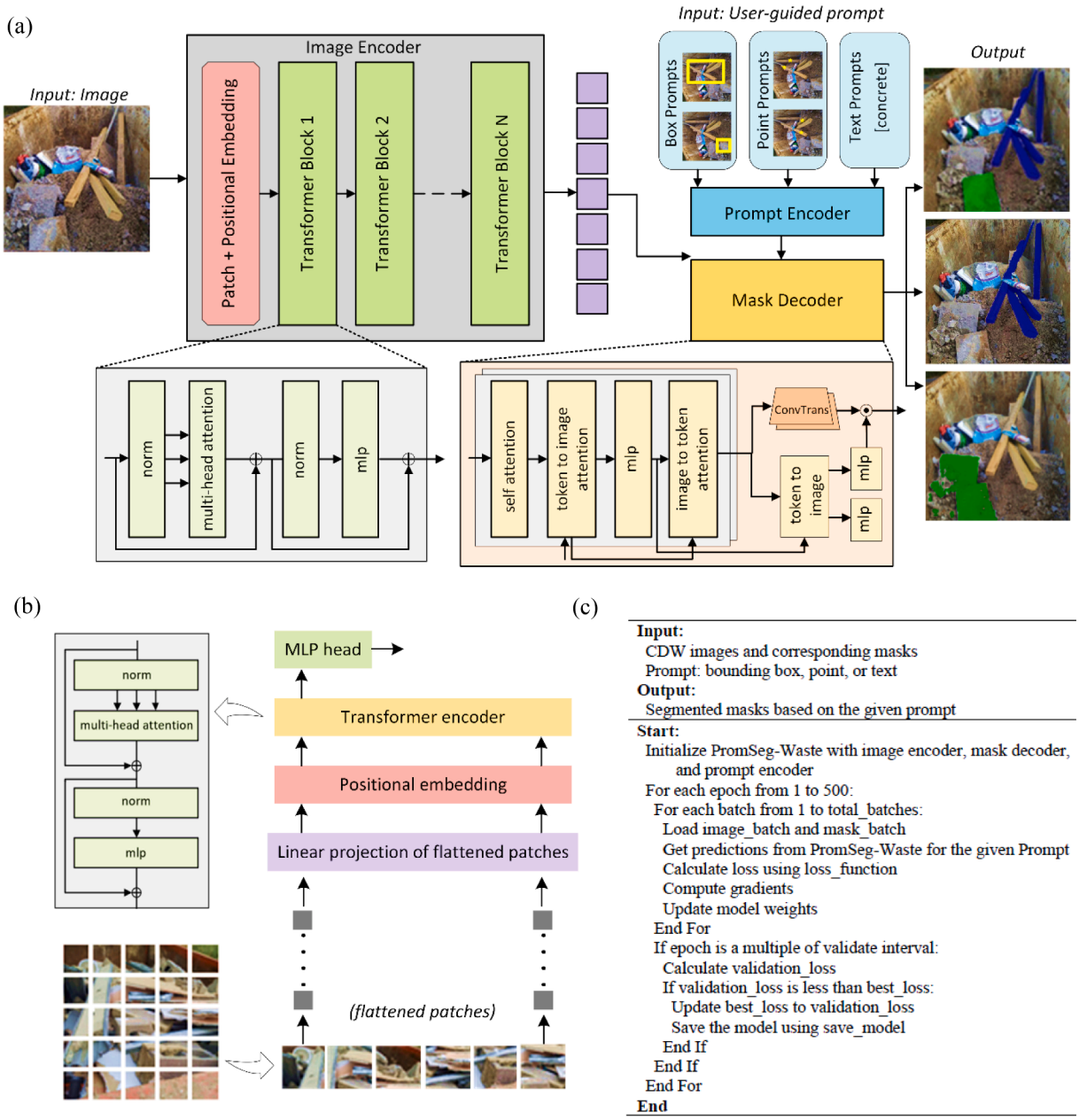


Fig. 2. (a) Architecture of PrompSeg-Waste: a promptable segmentation model for CDW segmentation, (b) Vision Transformer and (c) Pseudo code of training pipeline.

design allows the model to segment CDW in complex, real-world settings flexibly.

2.2. Training and evaluation pipeline

The PromSeg-Waste training process initiates loading and pre-processing a dataset of images and ground truth masks, applying data augmentation techniques, and feeding the pre-processed data into the model. Fig. 2(c) provides a pseudo-code that summarizes the training pipeline. The process involves optimizing a combination of Dice and cross-entropy losses using the AdamW optimizer. We use the Dice-Sørensen coefficient (DSC) (*Eq.* (1) (Li et al., 2019) and intersection over union (IoU) (*Eq.* (2) (Rahman and Wang, 2016) as performance metrics in the evaluation

$$DSC(A,B) = \frac{2|A \cap B|}{|A| + |B|} \tag{1}$$

$$IoU(A,B) = \frac{2|A \cap B|}{|A| + |B| - |A \cap B|} \tag{2}$$

(A and B are two sets of data: ground truth and prediction, where $0 \leq DSC, IoU(A,B) \leq 0$).

Bounding box prompt: The process starts by loading the ground truth mask and identifying the labels present in the ground truth. Then, a single label is chosen at random, and a binary mask is created for that label. The model then finds the coordinates of the non-zero pixels in this mask. The minimum and maximum x ($x_i min, x_i$max) and y ($y_i$min *and* $y_i$max) coordinates of these pixels define the initial bounding box. Random perturbations are added for variability, resulting in a slightly adjusted bounding box that captures the region of interest in the

mask as denoted in Eq. (3). The final bounding box ($bounding\ box_S$) coordinates are constrained within the image boundaries and used for training and evaluation.

$$boundingbox_S = [x_i min', y_i min', x_i max', y_i max'] \quad (3)$$

where, $x_i\min' = \max(0,\ x_i\min - rand(0,\ \alpha)), x_i\max' = \min(W,\ x_i\max + rand(0,\ \alpha)), y_i\min' = \max(0,\ y_i\min - rand(0,\ \alpha))$ and $y_i\max' = \min(H,\ y_i\max + rand(0,\ \alpha))$. Here, H and W are the height and width of the initial bounding box, α is the bounding box shift, and *rand* selects a random integer between 0 and α to shift the initial bounding box.

Point prompt: Similar to the bounding box method, a label is randomly selected, creating a binary mask. Then, the model determines the x and y coordinates ($x_i indices$ and $y_i indices$) of the non-zero pixels in this mask. Using these coordinates, a set of points is randomly sampled to act as point prompts, ensuring that these points are within the region of interest as defined in Eq. (4). These points are then fed into the prompt encoder.

$$coordinates = [x_i point, y_i point] \quad (4)$$

where, $x_i\,point = rand(x_i\,indices)\ and\ y_i\,point = rand(y_i\,indices)$. Here, *rand* selects random points from x and y indices.

Text prompt: The model is provided with images, ground truth masks, and text labels for each waste category as inputs. First, the input data is pre-processed, including data augmentation and text tokenization, using the CLIP (contrastive language-image pretraining) tokenizer (Radford et al., 2021). Then, a custom text prompt encoder that integrates a pre-trained CLIP text model processes the text. We define a dictionary containing label(s) corresponding to each class in the dataset (i.e. {class 2: 'concrete'}). Then, the model randomly selects a label, and the *CLIP_Tokenizer* converts it into a format suitable for the model as defined in Eq. (5). This text is transformed into embeddings, which are then processed by the *text_prompt_encoder* as in Eq. (6).

$$text\_token = CLIP\_Tokenizer\ ('concrete') \quad (5)$$

$$text\_embeddings = text\_prompt\_encoder\ (text\_token) \quad (6)$$

Model Predictions: For bounding box and point prompts, the model processes the coordinates and the image and encodes them into a lower-dimensional feature representation. The prompt encoder generates spatial context for mask prediction by creating sparse and dense embeddings from the coordinates. These are combined with image features and fed into the mask decoder to predict a low-resolution mask, which is then upscaled to the original image resolution using bilinear interpolation. For text prompts, the image is processed by an image encoder, while the text is tokenized and processed through the CLIP text encoder. A trainable encoder head projects the resulting text embeddings on the prompt to produce sparse prompt embeddings. The text and image embeddings are input into the mask decoder to make the final prediction, allowing accurate segmentation based on various prompts.

2.3. Construction and demolition waste (CDW) dataset

This study utilized the dataset gathered and prepared by Sirimewan et al. (2024a) for training and testing of the model. The dataset includes ten types of waste in construction and demolition sites captured from skip bins. We extracted 5277 image-mask pairs corresponding to ten distinct classes, as Fig. 3(a) described. Each image was manually annotated using an open-source tool to ensure accurate segmentation masks. The distribution of masks for each waste category is shown in

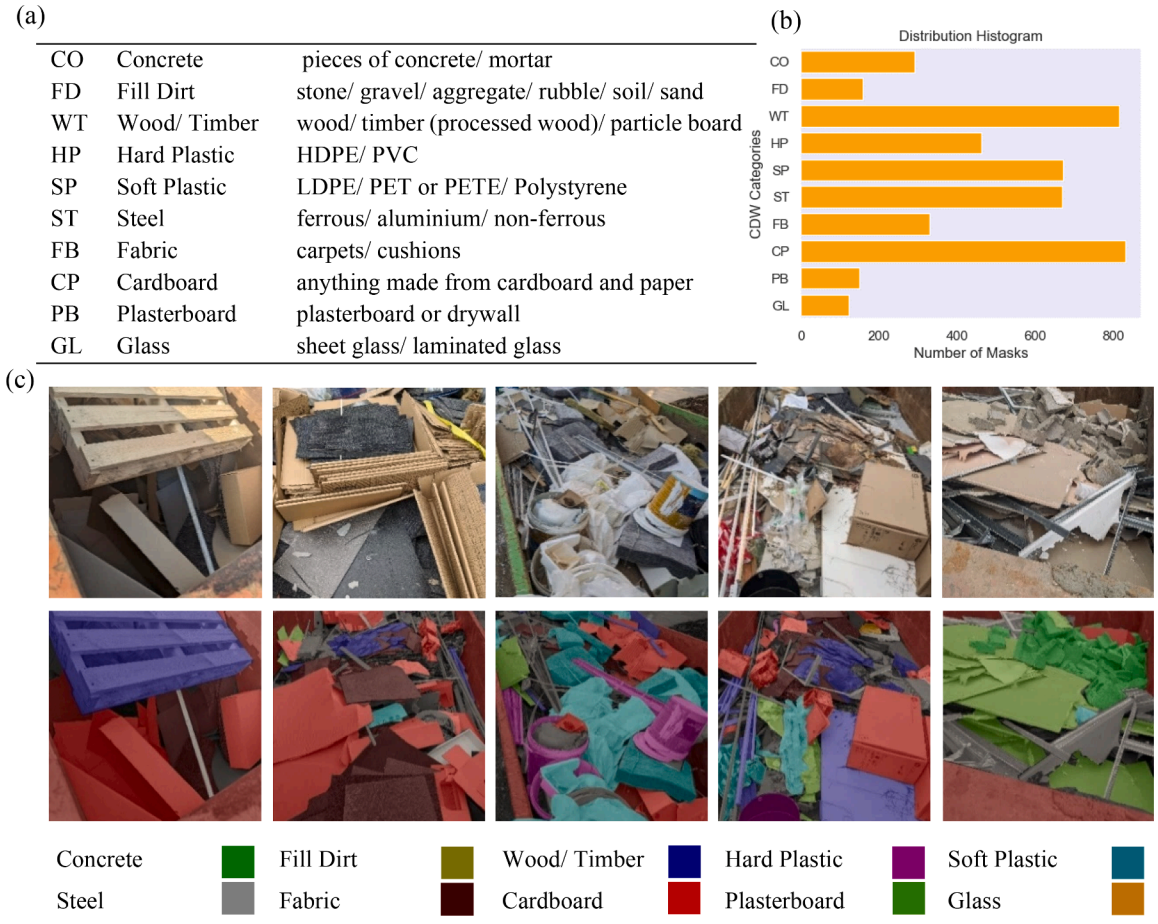


Fig. 3. (a) Descriptions of waste categories, (b) Distribution histogram over ten waste categories, and (c) Samples of image-mask pairs in the dataset.

Fig. 3(b). Fig. 3(c) displays some samples from the dataset, presenting the original images alongside their respective ground truth masks. These examples illustrate the diverse and complex nature of CDW in real-world environments, emphasising the need for robust segmentation models capable of handling such diversity.

3. Results and discussion

The PromSeg-Waste model is trained in a PyTorch environment utilizing an NVIDIA RTX A5500-24 GB GPU. Three distinct segmentation approaches are explored for bounding box, point, and text prompts. Previous studies suggest that scaling up ViT models, like ViT Large and ViT Huge, provided minimal gains in performance despite substantially higher computational requirements (Ma et al., 2024). Therefore, PromSeg-Waste utilizes the base ViT model as the image encoder to balance segmentation performance and computational efficiency.

3.1. User-guided bounding box prompt segmentation

The segmentation results using the bounding box prompt revealed satisfactory performance across various waste categories. The model achieves a mean Dice-Sørensen Coefficient (DSC) of 0.591 and a mean Intersection over Union (IoU) of 0.486, delineating ten waste categories from the test images. The concrete (CO) category demonstrates the highest performance among all waste categories, with a DSC of 0.7, highlighting its distinct and relatively uniform appearance in most instances. Despite the higher number of masks in the dataset, the timber (WT) and cardboard (CP) categories had slightly lower performances than concrete, with DSC scores of 0.66 and 0.63, respectively. This is possibly due to the similarities in visual features between timber and cardboard, making their differentiation challenging in some contexts. Hard plastic (HP), soft plastic (SP), and steel (ST) showed reasonably good performance, with DSC scores of around 0.6. However,

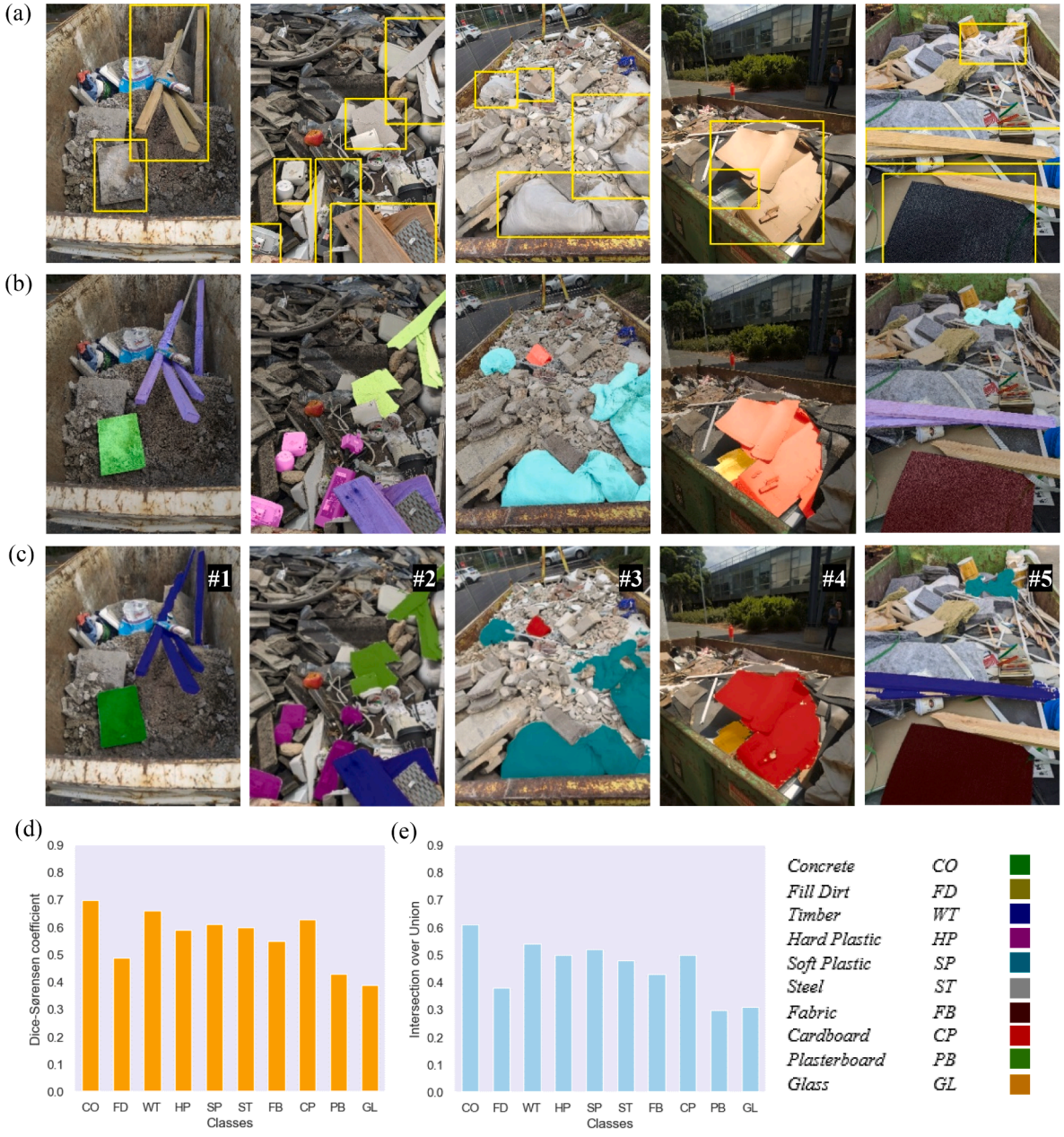


Fig. 4. Qualitative and quantitative results of **bounding box prompt-guided segmentation** of PromSeg-Waste model. (a) Source images, (b) Ground-truths, (c) Predictions of PromSeg-Waste, (d) and (e) Class-wise performance in terms of Dice-Sørensen coefficient and Intersection over Union.

plasterboard (PB) and glass (GL) classes had slightly below 50 % DSC scores. The lower performance can be attributed to the insufficient number of masks available for these classes during training, which limits the model's ability to learn and generalize effectively (Lu et al., 2022). When fewer examples exist, the model struggles to capture these materials' variability and distinct features, resulting in less accurate segmentation performance. Fig. 4 visually represents the qualitative and quantitative results of the PromSeg-Waste bounding box prompt-guided segmentation.

The qualitative results show that the bounding box prompt-guided training produced higher quality segmentations for most of the waste categories, such as concrete in (c)#1, hard plastic pieces in (c)#2, and timber and fabric pieces in (c)#5. However, the soft plastic in (c)#3 is not ideally segmented compared to the ground truth. This discrepancy may be due to the similar colours shared by the soft plastic and the white-painted concrete pieces in the image. Overall, the PromSeg-Waste model effectively segments waste categories based on bounding box prompts and provides quality segmentation masks with reasonable performance considering the complexity of the dataset. Its robust performance in automated waste recognition tasks highlights its practical utility for waste management applications. These include downstream processes such as waste monitoring, quality inspection and sorting at material recovery facilities (MRFs).

3.2. User-guided point prompt segmentation

The findings from the point prompt-guided segmentation highlight the difficulties of using single-point prompts in complex scenarios, as depicted in Fig. 5. Single-point prompts can be ambiguous because the model may not accurately determine which part of the object the user intends to segment (Kirillov et al., 2023). This limitation becomes evident in environments with overlapping objects or branch-type materials (Ma et al., 2024). We conducted an ablation study to assess the model's performance when given multiple points, as shown in Fig. 5. We aim to enhance the segmentation performance of the model by increasing the number of points. The results show that adding more points enhances the segmentation outcomes. However, we observe that beyond the fourth point, there is no significant improvement in performance.

The example shown in Fig. 5(a), #1 to #8, demonstrates how additional points can refine the segmentation task using point prompts. In particular, the #5 to #8 examples display the segmented areas of timber pieces using a single-point prompt. The prompt can be ambiguous in this example. Hence, the model faces difficulty recognizing them as a single

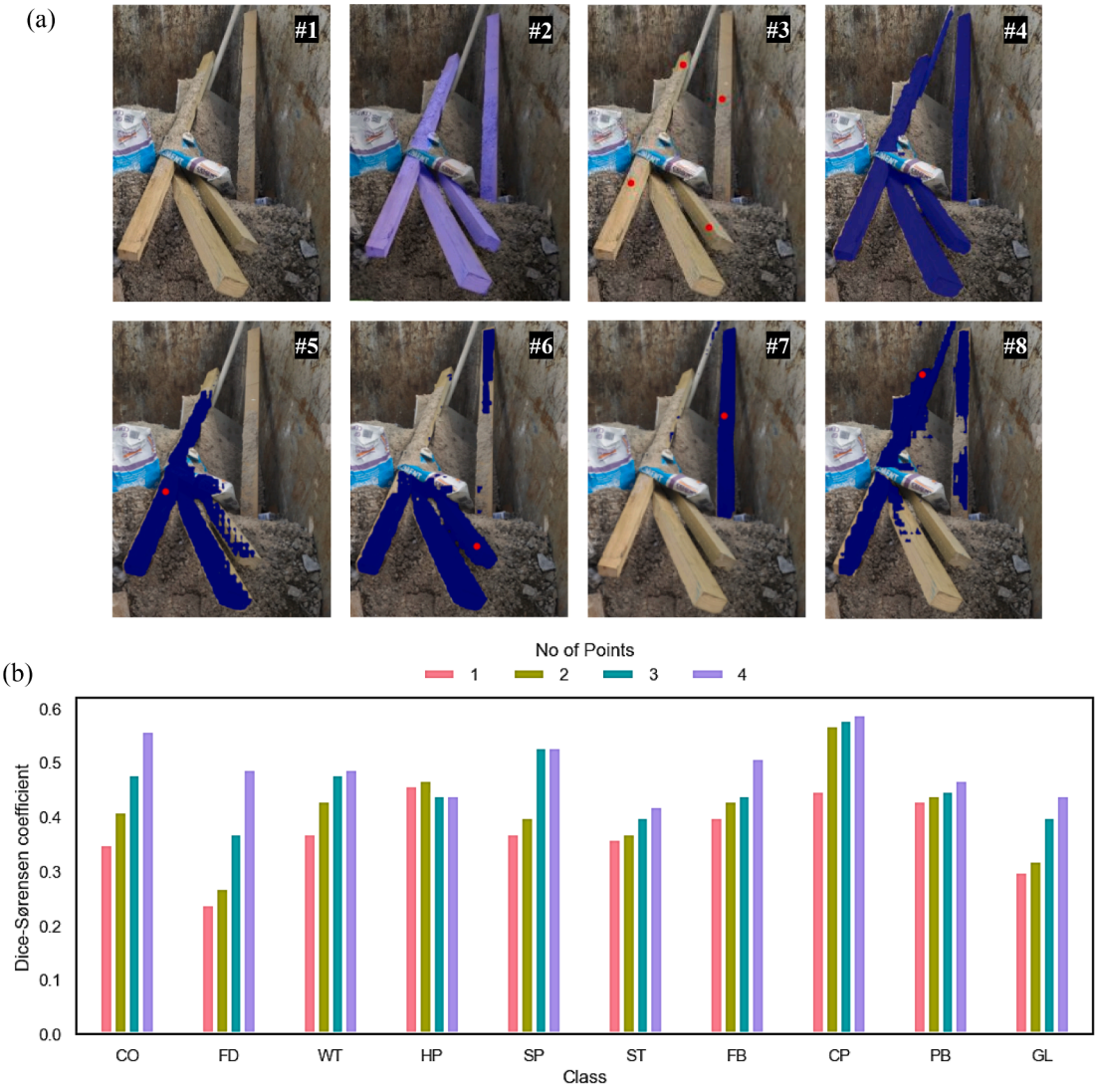


Fig. 5. Qualitative and quantitative results of **point prompt-guided segmentation** of PromSeg-Waste model. #1 – Source image, #2 – Ground-truth, #3 – Multiple point prompts, #4 – Prediction of multiple point prompts, #5 to #8 – Single point prompt and predictions, (b) Quantitative results of ablation study.

class. However, examples #3 and #4 demonstrate that using up to four points allows the model to identify and segment the target objects more precisely. This reduces the likelihood of errors that are prevalent with single-point prompts. The finetuned approaches, trained on domain-specific datasets, can learn the positional relationships of points within the target regions, thereby improving segmentation performance (Cheng et al., 2023).

Fig. 5(b) highlights the results of the ablation study, showing DSC scores across each waste category in the dataset. The model achieves a mean DSC score of 0.51 with four-point prompts. Regarding class-wise performance, concrete (CO), fill dirt (FD), timber (WT), soft plastic (SP), and cardboard (CP) categories exhibited DSC values of around 0.5. The fill dirt (FD) category performs poorly when only one to three points are used. This is due to the ambiguity in recognizing its extent, as it usually spreads over the entire bin. Hence, a single prompt is insufficient to recognize the whole area. In such cases, providing multiple prompts proves beneficial, as demonstrated by the results of the ablation study. Collectively, the findings of point prompt-guided segmentation illustrate that the PromSeg-Waste performs better with an increased number of points up to four. This iterative approach highlights the importance of user-guided input in refining segmentation quality in complex scenarios.

Bounding box and point prompt-guided segmentation methods facilitate the creation of accurately annotated datasets for model training and further research. These methods enable the model to detect and segment waste materials in cluttered environments. The users can choose diverse waste samples, including less common or challenging

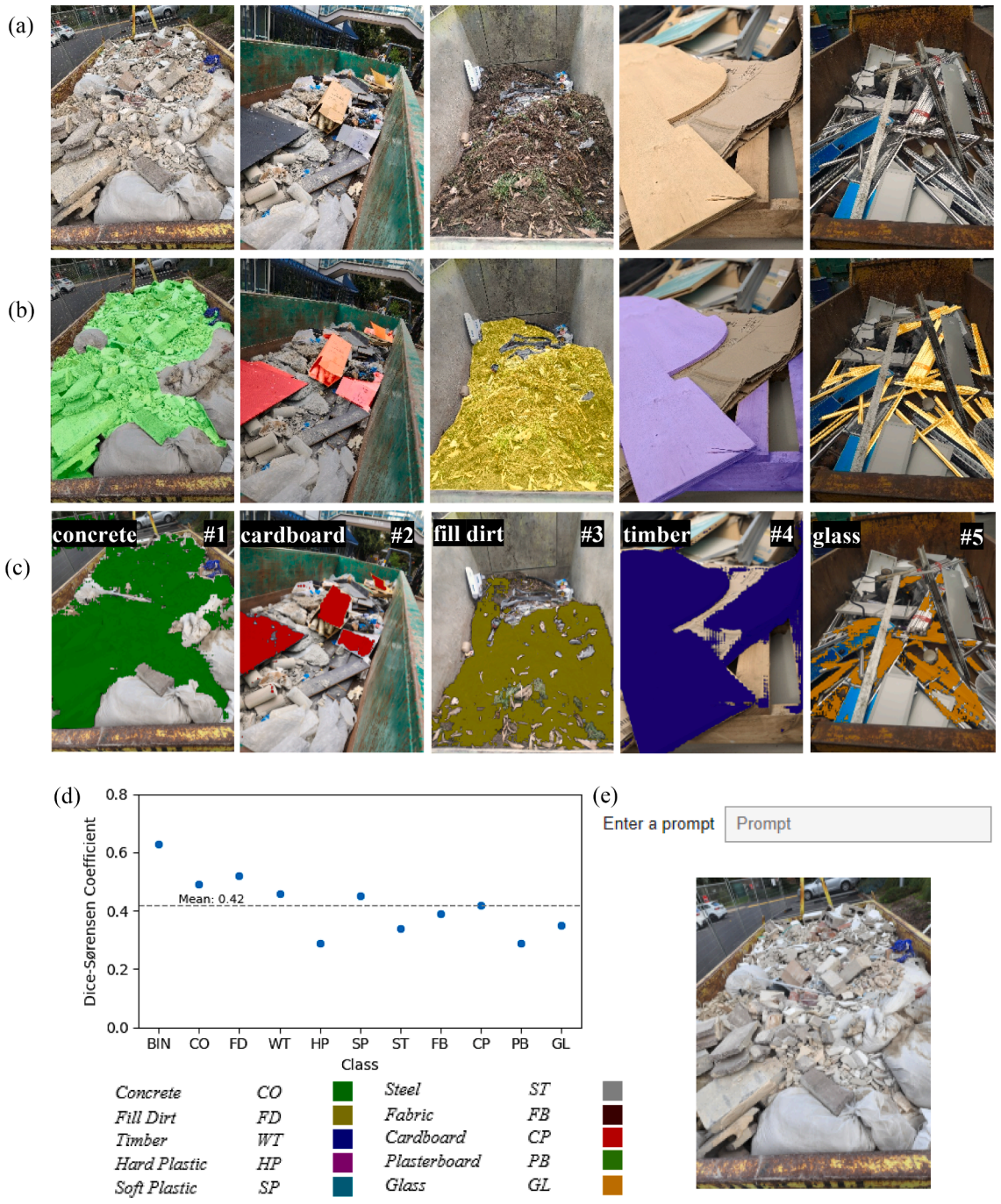


Fig. 6. Qualitative and quantitative results of **text prompt-guided segmentation** (a) Source images, (b) Ground-truths, (c) Predictions of PromSeg-Waste, (d) Dice-Sørensen coefficient for class-wise performance and (e) Example user-interface for PromSeg-Waste text prompt-guided segmentation.

types, to further improve the model's adaptability and performance across a broader range of scenarios. Moreover, this human–machine interface allows real-time adjustments and decision-making based on user inputs, enhancing the flexibility and precision of the segmentation process.

3.3. User-guided text prompt segmentation

The results of text prompt-guided segmentation demonstrate that PromSeg-Waste can generate segmentation masks for target waste objects using a simple text prompt. Using the class name as the label, we train the model to identify the ten types of waste materials, such as concrete, timber, etc. The results demonstrate that the model performs well across various classes, indicating its capability to segment waste materials based on text prompts. The model achieves a mean DSC of 0.42 across the classes, emphasising the need for further finetuning with sophisticated prompt engineering. The results of this study can serve as a baseline for future research on text prompt-guided segmentation. The qualitative and quantitative results of PromSeg-waste on text prompt-guided segmentation are presented in Fig. 6.

The qualitative results demonstrate that this method can effectively segment 'thing-type' materials, such as cardboard (#2), as well as 'stuff-type' materials, such as concrete (#1) and fill dirt (#3). This capability is beneficial for monitoring the quality of waste entering the MRFs. By segmenting the user's intended waste materials based on the text prompts, the model facilitates the assessment of waste load quality at the gates of MRFs. If a waste load contains a high amount of mixed or contaminated materials, it incurs higher processing costs and, hence, charges a higher gate fee. Conversely, waste loads with fewer or no mixed materials are less costly to process, leading to a lower gate fee (Tennakoon, 2024). This eliminates the need for direct manual inspection of incoming waste at MRFs. The proposed human–machine interface helps MRFs manage their operations more efficiently and maintain fair pricing based on the quality of the incoming loads with less manual involvement.

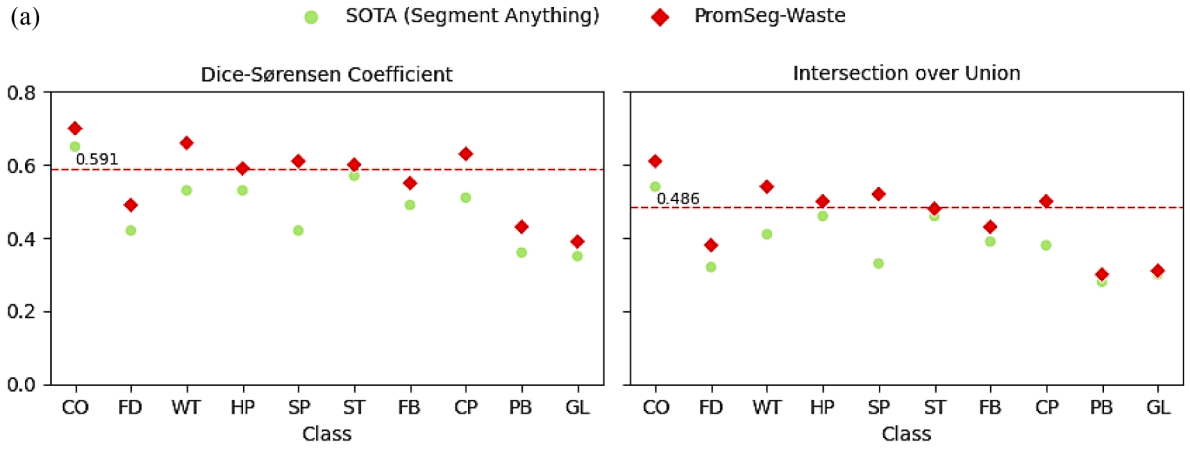
Example #4 in Fig. 6 shows the challenging nature of distinguishing between timber and cardboard due to their similar colours and textures. Similarly, #5 shows the model's difficulty in identifying glass mixed with other materials inside the bin. These examples emphasise the necessity for enhancing the model by including more samples from challenging classes, such as timber and cardboard, and underrepresented classes, such as glass and plasterboard, in the dataset. Improving the model with a more diverse and comprehensive training dataset will enhance its capability to identify and segment a broader range of materials in complex scenarios. Fig. 6(e) shows an example user interface for PromSeg-Waste text prompt-guided segmentation. The users can type the waste category name (i.e., concrete, timber) to generate a corresponding segmentation map, as shown in Fig. 6(c). This streamlines the interaction process, particularly when typing a category, which may be more efficient than providing spatial inputs like bounding boxes or points.

3.4. Comparison with the SOTA

The study compared the PromSeg-Waste bounding box prompt-guided segmentation results with a SOTA foundation model. The findings show that PromSeg-Waste has an increased performance of about 9 % within the context of this study. Fig. 7(a) presents comparisons of class-wise DSC and IoU and the mean values across the categories.

The most significant performance gap is observed in the soft plastic (SP) category, with a 19 % performance increase in the PromSeg-Waste model. This disparity can be attributed to the creased and deformed nature of soft plastics and the contamination commonly found in the context of CDW compared to generic images. Also, a higher performance improvement was observed in the timber (WT) and cardboard (CP) categories, with 13 % and 12 %, respectively, in PromSeg-Waste compared to SOTA. These categories exhibit more contaminations and deformations when derived from construction and demolition activities as waste, compared to the generic images of such materials. The minimal performance gaps were observed in the steel (ST) and glass (GL) categories, with only 3 % and 4 %, respectively, compared to SOTA. These can be attributed to the nature of these waste types in which those classes do not exhibit significant differences from generic image characteristics. The comparison emphasises that SOTA segmentation

(a)



(b)

SOTA pre-trained models	Total parameters (millions)	Training time (500 epochs)	Inference time (s)	Mean DSC (Bbox)
ViT Base	**93,735,472**	**6 hours**	**0.65**	**0.59**
ViT Large	312,342,832	17 hours	1.63	0.61
ViT Huge	641,090,608	31.5 hours	2.78	0.62

Fig. 7. (a) Comparison of PromSeg-Waste segmentation performance with the SOTA and (b) Comparison of model complexities and computational times.

methods are underperforming in downstream applications characterized by complex and challenging environments (Ma et al., 2024). In other words, these large-scale visual foundation models require further refinement, especially for specific applications in practical contexts (Xiong et al., 2024). This comparison highlights the necessity for improvements in SOTA promptable segmentation models to enhance their usefulness in specialized real-world applications.

In addition, the study compared the computational demands for model training and inference with two other SOTA ViT models. The model complexities and the computational times required for the experiments on an NVIDIA RTX A5500-24 GB GPU setting are shown in Fig. 7(b). The base ViT model performed segmentation with a computation time of 0.65 s per image, which is significantly lower than larger models like ViT Large and ViT Huge. For example, ViT Huge required 2.78 s per image, translating to a four times increase in inference time while only offering marginal gains in segmentation performance. This highlights the viability of using the base model in applications with limited computational resources.

3.5. Generalizability and interpretability of PromSeg-Waste

The CDW environment is diverse and complex, with waste materials varying depending on the site, weather conditions, and stages of construction and demolition (Davis et al., 2021). In practical applications, the CDW encountered at a residential building may differ considerably from that of a large-scale commercial or industrial demolition. Recognizing these variations, the PromSeg-Waste is developed with the flexibility to handle mixed waste streams by leveraging user-guided prompts to adapt to different scenarios. Although this study demonstrated the model's effectiveness on a specific dataset, its flexible prompting mechanism suggests it could generalize well across various environments by adjusting prompts for specific waste types or site conditions. For example, in particularly cluttered scenes, the model can be provided with multiple user-guided prompts to improve its segmentation performance. The ablation study demonstrated that giving additional prompts helped refine the segmentation of waste categories. Furthermore, the model's scalability is supported by its adaptable architecture, which can be finetuned on more extensive and more diverse datasets. This improves its ability to perform across varied environments, including different waste handling facilities or geographic regions.

Interpretability is critical in practical applications of deep learning models, where trust in the model's outputs is essential for decision-making (Chakraborty et al., 2017). To ensure transparency in the decision process, several mechanisms are incorporated in PromSeg-Waste that make its predictions more interpretable to the end user. One key feature is the use of user-guided prompts, such as bounding boxes, points, and text, which directly involve the user in guiding the model's segmentation. This allows users to understand how their inputs influence the segmentation results, making the process more transparent and giving them control over adjustments. Additionally, the model's outputs are visualized as segmentation maps, delineating the boundaries of different waste categories based on the input prompts. Users can interactively refine these outputs, and the changes are reflected in real-time. This offers a clear connection between input adjustments and model predictions. This iterative process provides users with insights into how the model interprets different types of waste based on their prompts, further enhancing trust.

4. Limitations and future work

Some classes, including plasterboard, glass, and fabric, have fewer masks than others. This makes it challenging for the model to recognize these underrepresented classes in cluttered settings, impacting overall performance. Hence, addressing the data imbalance is essential by including more samples from the challenging classes to improve performance. Future work can expand on this by evaluating the model with a more comprehensive set of waste categories and identifying contaminants in recycled waste, contributing to the quality assurance of recycled products. Certain complexities may arise in the practical application at MRFs. Providing precise prompts may become challenging for the user in highly cluttered environments or when waste materials are overlapping. To address this challenge, further studies can be focused on incorporating a feedback mechanism in the system, allowing users to iteratively refine their prompts based on visual feedback from the model's predictions. Challenges to the model's generalizability may arise in extreme conditions, such as low-light environments, weather interference, or severely cluttered waste streams. In these cases, the model's adaptability allows for incremental improvements by finetuning the model with diverse datasets that reflect such conditions. On the other hand, domain adaptation will enable the model to finetune its parameters when faced with entirely new waste streams or environments. Also, future work could explore real-world scalability by deploying the model in MRFs to validate its robustness and efficiency in diverse, practical conditions.

5. Conclusion

This study addresses the research gap in prompt-guided segmentation for CDW, an overlooked domain, despite its significance for improving waste management operations. Utilizing a real-world dataset comprised of ten distinct CDW categories, we develop the 'PromSeg-Waste' pipeline to segment waste materials through user-guided prompts such as bounding boxes, points, and text. We design this model with practical usability in mind. Simple and natural prompts make the system accessible to non-experts working in MRFs. For instance, bounding boxes and points can quickly be drawn using a mouse or touchscreen on a user-friendly interface. This allows the operators to interact with the system without needing advanced training in machine learning or segmentation techniques.

The model achieves a mean DSC of 0.591 for bounding box prompt-guided segmentation across ten waste categories. This reflects a notable 9 % improvement over the existing SOTA. The findings of this study establish a baseline for prompt-guided segmentation in the context of CDW. PromSeg-Waste offers a flexible prompting system capable of real-time interaction, making it suitable for dynamic environments where quick decision-making is required. Additionally, the model enhances transparency and user trust by allowing interactive feedback that refine segmentation performance. This dedicated interactive segmentation method can integrate with workflows at MRFs as a human–machine interface to minimize the worker's exposure to waste handling processes. It can enhance the efficiency of MRFs by creating a more cost-effective and time-saving process with minimal labor involvement in waste monitoring and sorting. Moreover, the approach offers an efficient semi-automated approach to data annotation, creating tailored datasets for downstream tasks. These practical applications highlight the model's potential to streamline the waste handling process, offering time and resource savings while minimizing human exposure to waste.

CRediT authorship contribution statement

Diani Sirimewan: Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Nilakshan Kunananthaseelan:** Writing – review & editing, Methodology. **Sudharshan Raman:** Writing – review & editing, Project administration, Funding acquisition, Conceptualization. **Reyes Garcia:** Writing – review & editing, Project administration. **Mehrdad Arashpour:** Writing – review & editing, Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization.

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.

Data availability

Data will be made available on request.

Acknowledgement

The authors are also grateful for support from the ASCII Lab (https://www.monash.edu/ascii) members at Monash University and their constructive feedback on progressive iterations of this work.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.wasman.2024.09.018.

References

Arashpour, M., 2023. AI explainability framework for environmental management research. J. Environ. Manage. 342, 118149.
Chakraborty, S., Tomsett, R., Raghavendra, R., Harborne, D., Alzantot, M., Cerutti, F., Srivastava, M., Preece, A., Julier, S., Rao, R. M., 2017. Interpretability of deep learning models: A survey of results. 2017 IEEE smartworld, ubiquitous intelligence & computing, advanced & trusted computed, scalable computing & communications, cloud & big data computing, Internet of people and smart city innovation (smartworld/SCALCOM/UIC/ATC/CBDcom/IOP/SCI).
Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European conference on computer vision (ECCV).
Chen, J., Fu, Y., Lu, W., Pan, Y., 2023. Augmented reality-enabled human-robot collaboration to balance construction waste sorting efficiency and occupational safety and health. J. Environ. Manage. 348, 119341.
Chen, X., Huang, H., Liu, Y., Li, J., Liu, M., 2022. Robot for automatic waste sorting on construction sites. Autom. Constr. 141, 104387.
Cheng, J., Ye, J., Deng, Z., Chen, J., Li, T., Wang, H., Su, Y., Huang, Z., Chen, J., Jiang, L., 2023. Sam-med2d. *arXiv preprint arXiv:2308.16184*.
Chileshe, N., Jayasinghe, R.S., Rameezdeen, R., 2019. Information flow-centric approach for reverse logistics supply chains. Autom. Constr. 106, 102858.
Davis, P., Aziz, F., Newaz, M.T., Sher, W., Simon, L., 2021. The classification of construction waste material using a deep convolutional neural network. Autom. Constr. 122. https://doi.org/10.1016/j.autcon.2020.103481.
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 Manag. 167, 194–203.
Demetriou, D., Mavromatidis, P., Petrou, M.F., Nicolaides, D., 2024. CODD: A benchmark dataset for the automated sorting of construction and demolition waste. Waste Manag. 178, 35–45.
Di Maria, F., Bianconi, F., Micale, C., Baglioni, S., Marionni, M., 2016. Quality assessment for recycling aggregates from construction and demolition waste: An image-based approach for particle size estimation. Waste Manag. 48, 344–352.
Dodampegama, S., Hou, L., Asadi, E., Zhang, G., Setunge, S., 2024. Revolutionizing construction and demolition waste sorting: Insights from artificial intelligence and robotic applications. Resour. Conserv. Recycl. 202, 107375.
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.
Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.
Fu, Y., Chen, J., Lu, W., 2024. Human-robot collaboration for modular construction manufacturing: review of academic research. Autom. Constr. 158, 105196.
He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision.
Islam, R., Nazifa, T.H., Yuniarto, A., Uddin, A.S., Salmiati, S., Shahid, S., 2019. An empirical study of construction and demolition waste generation and implication of recycling. Waste Manag. 95, 10–21.
Jayasinghe, R.S., Chileshe, N., Rameezdeen, R., 2019. Information-based quality management in reverse logistics supply chain: a systematic literature review. BIJ 26 (7), 2146–2187.
Jin, S., Yang, Z., Królczykg, G., Liu, X., Gardoni, P., Li, Z., 2023. Garbage detection and classification using a new deep learning-based machine vision system as a tool for sustainable waste recycling. Waste Manag. 162, 123–130.
Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., 2023. Segment anything. In: Proceedings of the IEEE/CVF International Conference on Computer Vision.
Laadila, M.A., LeBihan, Y., Caron, R.-F., Vaneeckhaute, C., 2021. Construction, renovation and demolition (CRD) wastes contaminated by gypsum residues: characterization, treatment and valorization. Waste Manag. 120, 125–135.
Li, Z., Deng, Q., Liu, P., Bai, J., Gong, Y., Yang, Q., Ning, J., 2024. An intelligent identification and classification system of decoration waste based on deep learning model. Waste Manag. 174, 462–475.
Li, X., Sun, X., Meng, Y., Liang, J., Wu, F., Li, J., 2019. Dice loss for data-imbalanced NLP tasks. *arXiv preprint arXiv:1911.02855*.
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.
Li, Y., Zhang, X., 2024. Multi-modal deep learning networks for RGB-D pavement waste detection and recognition. Waste Manag. 177, 125–134.
Lu, W., Chen, J., Xue, F., 2022. Using computer vision to recognize composition of construction waste mixtures: a semantic segmentation approach. Resour. Conserv. Recycl. 178, 106022.
Lu, W., Chen, J., 2022. Computer vision for solid waste sorting: a critical review of academic research. Waste Manag. 142, 29–43. https://doi.org/10.1016/j.wasman.2022.02.009.
Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B., 2024. Segment anything in medical images. Nat. Commun. 15 (1), 654.
Majchrowska, S., Mikołajczyk, A., Ferlin, M., Klawikowska, Z., Plantykow, M.A., Kwasigroch, A., Majek, K., 2022. Deep learning-based waste detection in natural and urban environments. Waste Manag. 138, 274–284.
Mazurowski, M.A., Dong, H., Gu, H., Yang, J., Konz, N., Zhang, Y., 2023. Segment anything model for medical image analysis: an experimental study. Med. Image Anal. 89, 102918.
Moral, P., García-Martín, Á., Escudero-Viñolo, M., Martínez, J.M., Bescós, J., Peñuela, J., Martínez, J.C., Alvis, G., 2022. Towards automatic waste containers management in cities via computer vision: containers localization and geo-positioning in city maps. Waste Manag. 152, 59–68.
Na, S., Heo, S., Han, S., Shin, Y., Lee, M., 2022. Development of an artificial intelligence model to recognise construction waste by applying image data augmentation and transfer learning. Buildings 12 (2), 175.
Nežerka, V., Zbíral, T., Trejbal, J., 2024. Machine-learning-assisted classification of construction and demolition waste fragments using computer vision: convolution versus extraction of selected features. Expert Syst. Appl. 238, 121568.
Pitakaso, R., Srichok, T., Khonjun, S., Golinska-Dawson, P., Gonwirat, S., Nanthasamroeng, N., Boonmee, C., Jirasirilerd, G., Luesak, P., 2024. Artificial Intelligence in enhancing sustainable practices for infectious municipal waste classification. Waste Manag. 183, 87–100.
Prasad, V., Arashpour, M., 2024. Optimally leveraging depth features to enhance segmentation of recyclables from cluttered construction and demolition waste streams. J. Environ. Manage. 354, 120313.
Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., 2021. Learning transferable visual models from natural language supervision. International conference on machine learning.
Rahman, M. A., Wang, Y., 2016. Optimizing intersection-over-union in deep neural networks for image segmentation. International symposium on visual computing.
Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18.
Sirimewan, D., Bazli, M., Raman, S., Mohandes, S.R., Kineber, A.F., Arashpour, M., 2024a. Deep learning-based models for environmental management: recognizing construction, renovation, and demolition waste in-the-wild. J. Environ. Manage. 351, 119908.
Sirimewan, D., Harandi, M., Peiris, H., Arashpour, M., 2024b. Semi-supervised segmentation for construction and demolition waste recognition in-the-wild: adversarial dual-view networks. Resour. Conserv. Recycl. 202, 107399.
Tennakoon, G., Rameezdeen, R., Chileshe, N., 2022. Diverting demolition waste toward secondary markets through integrated reverse logistics supply chains: a systematic literature review. Waste Manag. Res. 40 (3), 274–293.
Tennakoon, G. A., 2024. *Towards Circularity in Construction: Promoting the Uptake of Reprocessed Construction Materials* University of South Australia]. https://find.library.unisa.edu.au/discovery/delivery/61USOUTHAUS_INST:ROR/12285929750001831.
Véliz, K., Ramírez-Rodríguez, G., Ossio, F., 2022. Willingness to pay for construction and demolition waste from buildings in Chile. Waste Manag. 137, 222–230.
Wang, X., Han, W., Mo, S., Cai, T., Gong, Y., Li, Y., Zhu, Z., 2023. Transformer-based automated segmentation of recycling materials for semantic understanding in construction. Autom. Constr. 154, 104983.
Wei, Z., Chen, P., Yu, X., Li, G., Jiao, J., Han, Z., 2024. Semantic-aware SAM for Point-Prompted Instance Segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.
Wu, J., Xu, M., 2024. One-prompt to segment all medical images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.
Wu, X., Kroell, N., Greiff, K., 2024. Deep learning-based instance segmentation on 3D laser triangulation data for inline monitoring of particle size distributions in construction and demolition waste recycling. Resour. Conserv. Recycl. 205, 107541.
Xiong, Y., Varadarajan, B., Wu, L., Xiang, X., Xiao, F., Zhu, C., Dai, X., Wang, D., Sun, F., Iandola, F., 2024. Efficientsam: Leveraged masked image pretraining for efficient segment anything. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.

Yazdani, M., Kabirifar, K., Frimpong, B.E., Shariati, M., Mirmozaffari, M., Boskabadi, A., 2021. Improving construction and demolition waste collection service in an urban area using a simheuristic approach: a case study in Sydney, Australia. *J. Clean. Prod.* 280, 124138.

Yong, Q., Wu, H., Wang, J., Chen, R., Yu, B., Zuo, J., Du, L., 2023. Automatic identification of illegal construction and demolition waste landfills: a computer vision approach. Waste Manag. 172, 267–277.

Zou, X., Yang, J., Zhang, H., Li, F., Li, L., Wang, J., Wang, L., Gao, J., Lee, Y. J., 2024. Segment everything everywhere all at once. *Advances in neural information processing systems*, *36*.