# A Computer Vision Approach for Monitoring Construction Waste Deposition Events in Static Dumpsters

Benjamin Rafatian, benjamin.rafatian.2@ens.etsmtl.ca
*Department of Construction, École de technologie supérieure, Canada*

Ali Motamedi, ali.motamedi@etsmtl.ca
*Department of Construction, École de technologie supérieure, Canada*

Younes Driouache, younes.driouache@altaroad.com
*Altaroad, France*

Tarik Fahlaoui, tarik.fahlaoui@altaroad.com
*Altaroad, France*

**Abstract**
Although advanced technologies are increasingly used in the construction industry, there remains a gap in automating the monitoring of construction and demolition waste. This research investigates automating the monitoring of dumping events in waste management. Focusing on the initial phase of a complete monitoring framework, the study proposes a system employing computer vision algorithms to detect waste dumping in static dumpsters on construction sites or waste sorting facilities. The developed system eliminates the need for manual monitoring, reducing costs and increasing efficiency. Data collection, cleaning, and annotation efforts were made to curate a high-quality dataset using real-world data. This foundational work sets the stage for future advancements, such as integrating the system with material detection models and volume estimation techniques to enable precise identification and quantification of waste types, resulting in a complete monitoring framework for construction and demolition waste management.

**Keywords:** Computer Vision, Construction and Demolition Waste, Supply Chain Visibility

## 1 Introduction

The construction industry has witnessed remarkable advancements in the use of technology over the past few decades. Innovations such as Building Information Modeling (BIM), robotics, and the Internet of Things (IoT) have revolutionized construction practices, leading to improved efficiency, safety, and productivity. However, these advancements have also brought to light significant environmental challenges, particularly related to the management of construction and demolition waste (CDW).

CDW is a major contributor to environmental degradation. Approximately 35% of CDW generated worldwide is estimated to be disposed of in landfills without undergoing any further processing (Menegaki and Damigos, 2018). Inefficient waste management practices can lead to various issues, such as increased pollution, resource depletion, and negative impacts on human health. Despite the technological advancements in construction processes, waste management practices have not kept pace, leading to a critical gap in the industry's ability to effectively handle waste.

The primary problem addressed in this research is the need for an automated system to monitor and manage the disposal of waste materials in construction environments. Traditional methods of waste monitoring rely heavily on manual inspection and reporting, which are labor-intensive, error-prone, and often inadequate for real-time decision-making. This gap in waste management practices underscores the necessity for innovative solutions that leverage advanced technologies to enhance efficiency and accuracy.

To address these challenges, this research proposes a three-step framework for monitoring waste disposal in construction dumpsters (i.e., event detection, material recognition, and volume estimation). The framework aims to facilitate the monitoring process, from identifying waste disposal events to classifying materials and estimating the volume of generated waste.

The scope of this paper is to report on the developed event detection module of this framework. Event detection focuses on identifying instances where waste is deposited into dumpsters on construction sites. This involves detecting the approach and action of workers or vehicles carrying waste towards the dumpsters, using advanced computer vision techniques. By accurately detecting these events, the system can provide real-time monitoring and data collection, paving the way for more efficient and sustainable waste management practices in the construction industry.

By addressing the first step of the proposed framework, this paper presents the foundation for a comprehensive solution to the challenges of construction waste management. The subsequent steps of material recognition and volume estimation are essential to complete the system's capabilities, ultimately contributing to a more environmentally responsible construction industry.

## 2 Background Work

### Construction and Demolition Waste (CDW) Management

CDW management has become a critical focus in the pursuit of sustainable construction practices. Various studies have highlighted the importance of effective CDW management strategies to mitigate environmental impacts. Gálvez-Martos et al., (2018) provide a comprehensive overview of best management practices for CDW in Europe, emphasizing the role of regulatory frameworks and innovative technologies in improving waste management outcomes. Hossain et al., (2017) conduct a comparative environmental evaluation of different CDW management systems, highlighting the need for efficient sorting and recycling strategies to minimize environmental impact.

### Advances in Waste Management Technologies

Advancements in technology have significantly impacted the efficiency and effectiveness of waste management practices. For example, Bonifazi et al., (2019) discussed the use of hyperspectral imaging (HSI) for material identification in waste recycling processes, demonstrating the integration of HSI with machine learning techniques to enhance material characterization. Serranti et al., (2015) also explored the application of HSI for demolition waste recycling, emphasizing its potential for improving the quality control of recycled aggregates.

### Computer Vision for Waste Sorting

The application of computer vision (CV) in waste sorting has shown significant promise in recent years. Lu and Chen (2022) review the state-of-the-art in CV applications for solid waste sorting, identifying key algorithms and methodologies that have been developed to enhance sorting accuracy and efficiency. They highlight the shift from traditional machine learning methods, which rely on handcrafted features, to deep learning models that automatically learn features from data, significantly improving robustness and accuracy.

Wang et al., (2020) present a vision-based robotic system for on-site CDW sorting and recycling, demonstrating the integration of simultaneous localization and mapping (SLAM) and CV technologies to improve the efficiency of waste sorting operations. Similarly, Dong et al., (2022) introduce an innovative boundary-aware transformer model designed to identify the composition of construction waste. This model utilizes transformer-based semantic segmentation to accurately recognize waste materials at a detailed level.

### Object Detection Using Deep Learning

Deep learning models have revolutionized object detection tasks, enabling more accurate and efficient identification of objects in various contexts. The YOLO (You Only Look Once) family of models, introduced by Redmon et al., (2016), has become particularly popular for real-time object detection because of its speed and accuracy. The latest iteration at the time of this project, YOLOv8, offers several model sizes that balance inference speed and accuracy, making it suitable for deployment in resource-constrained environments.

Despite the advancements in object detection using deep learning, previous research has not focused on developing a comprehensive waste monitoring platform, specifically for detecting dumping events into static dumpsters on construction sites. While much of the construction and demolition waste (CDW) is managed through various means, including recycling facilities and landfill, a significant portion is initially disposed of in dumpsters. Most existing studies have concentrated on general waste classification and recycling processes but have not addressed the entire monitoring process for static dumpsters, from event detection to material recognition and volume estimation. This gap in the literature highlights the need for a specialized system that automates the monitoring of construction waste disposal events into dumpsters, ensuring more efficient and accurate waste management practices.

## 3 Methodology

This research project is designed to automate the monitoring of waste material disposal into the dumpsters in construction environments. The key objective is to develop a method to address the initial step of full waste monitoring framework (i.e., event detection) for static dumpsters. The developed solution should be able to detect any event of dumping waste material into the static dumpsters located on the site.

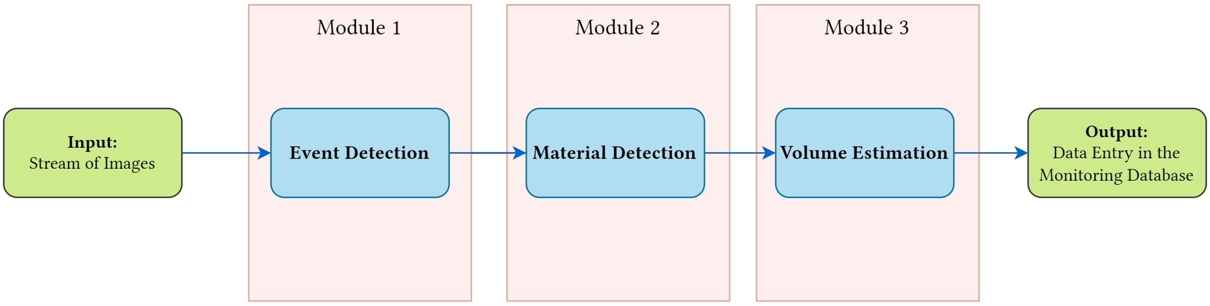


**Figure 1.** Full waste monitoring framework.

Primary goal of the Event Detection system is to identify the events of waste material being dumped into stationary dumpsters. In this project, a dumping event is defined as the approach of a human (worker on the site) or a vehicle (e.g., mini loader, excavator, loader, lift truck.) carrying waste material towards the dumpster and depositing it. This activity is recognized by first locating the dumpster in the field of view (FoV) of the camera, identifying humans or vehicles, tracking them, and then determining whether they are approaching or moving away from the dumpster based on their relative movement.

In this study, two approaches were evaluated to address this issue: the background subtraction method and object detection using deep learning models.

### 3.1 Background Subtraction

Background subtraction is an image processing technique that segments objects of interest in an image. As the camera's viewpoint is fixed and the dumpster remains stationary, this technique can be used to identify and locate moving objects within the camera's field of view. These moving objects, which are the humans or vehicles of interest, are potentially carrying waste materials to be dumped into the dumpster. By removing the static background and creating a foreground mask that represents the human or vehicle as a white blob, the movement of this white blob can be tracked to determine whether it is approaching the dumpster.

This method does not require training, as it relies solely on the camera feed and compares consecutive frames to identify moving elements as foreground and static elements as background. One of its drawbacks is the need for manual positioning of the dumpster in the image for each setup, as the method cannot identify the dumpster, considering it part of the background since it does not move across consecutive frames. Additionally, there are other limitations such as the accuracy, and the inability to detect the activity or details of the event, which are discussed subsequently.

To eliminate the background and identify the moving objects the `MOG2` algorithm (Zivkovic, 2004) was applied. A series of preprocessing steps were performed to prepare the raw image for the `MOG2` algorithm. The initial step involved changing the color space using OpenCV's `cvtColor` function, converting the original RGB image to a single-channel grayscale image. Subsequently, to smooth out sharp edges in the grayscale image, a Gaussian blur with a kernel size of $5 \times 5$ was applied.

The MOG2 background subtractor in OpenCV was configured with three key parameters: varThreshold, History, and DetectShadows. The varThreshold parameter sets the threshold for the squared Mahalanobis distance between the pixel and the model, which determines if a specific pixel is adequately represented by the background model(Theophanous, 2020). The History parameter specifies the number of last frames that affect the background model, essentially determining how quickly the model adapts to changes in the scene. The DetectShadows parameter enables the detection of shadows, which can be useful for improving the accuracy of object detection.

To apply the background subtractor to the processed image, the learning rate parameter, alpha, was calculated dynamically. The learning rate controls how quickly the background model updates, with a higher rate allowing for faster adaptation to changes but potentially increasing noise. The learning rate was variably calculated based on equation (1), which adjusts alpha according to the standard deviation of the single-channel grayscale image and the varThreshold parameter set for the MOG2 background subtractor.

$$\alpha = 0.001 + (1 - 0.001) \times \left(\frac{std}{varThreshold}\right) \quad (1)$$

This process produces a foreground mask. Binary thresholding is then applied to convert the foreground mask image into a binary image, where pixels representing the moving objects are rendered white, while pixels representing the background are rendered black. The approach of dynamically adjusting the learning rate is inspired by adaptive background subtraction methods, such as the foundational work by Zivkovic (2004), which suggest tuning parameters based on the variance in the image to improve the model's responsiveness to changes in the scene.

Next contours are drawn to locate the objects on the processed binary image, utilizing the original algorithm developed by Suzuki and Be (1985). Consequently, the area of each contoured object in each frame is calculated. If the area exceeds a set threshold (determined manually based on the settings and relative size of the objects in the image), the contoured object is considered as a human or vehicle. The logic then compares the center point of the detected human or vehicle to the predefined dumping zone. If the center point moves into the defined dumping zone and remains there for a sufficient duration, it is identified as a dumping event.

This method had multiple limitations. The most significant issue was that it could not accurately determine whether a human or vehicle was approaching the dumpster, relying solely on the movement of these objects, and was vulnerable to detecting other moving objects not intended as targets. Another limitation was its susceptibility to varying lighting conditions and harsh environments. To address this issue, efforts were made to convert the original RGB image into different color spaces and compare the result with the grayscale image. HSV, LAB, YUV, LUV, and XYZ color spaces were tested. In some challenging cases, such as images taken at night or during rain, some of these color spaces showed improvement over the grayscale image. However, this approach did not achieve the desired accuracy, as it still struggled with inconsistent performance and sensitivity to environmental variations.

Therefore, the second methodology was explored, involving object detection using a deep learning model. Although this approach was more complex to implement, it provided more opportunities and addressed some of the issues of the first method.

### 3.2 Object Detection Using Deep Learning

Object detection is a computer vision technique used to identify instances of objects in images or videos. In this research project, the objective is to identify instances of objects in three classes: dumpster, human, and vehicle. Various deep learning models are available for object detection. The YOLO family was chosen due to its performance, ease of implementation, real-time inference capabilities, and availability of models in various sizes. Since its initial release in 2016, several versions have been introduced by researchers in the field. In this project, YOLOv8 by Ultralytics was chosen, as it was the latest version at the time of implementation.

YOLOv8 offers five model sizes, ranging from nano to x-large, differing in the structure of their neural networks. The smallest model, YOLOv8n, has 3.2 million parameters, while the largest, YOLOv8x, has 68.2 million parameters. YOLO models perform various computer vision tasks, such as object detection, object segmentation, classification, and pose estimation. YOLO object detection models are pretrained on the COCO (Common Objects in Context) dataset (Lin et al., 2015), a large-scale image dataset with over 330,000 images and 1.5 million object instances across 80 different classes. A significant advantage of YOLO models is the ability to export the trained model to ONNX format, which is suitable for production environments.

### 3.2.1 Data Collection and Dataset Creation

Although the COCO dataset is extensive, it does not contain the specific classes of objects required for this project. The COCO dataset includes general vehicle categories, such as bicycle, car, motorcycle, truck, airplane, bus, and train, but it does not include the specific types of industrial vehicles we are targeting, such as lift trucks and excavators. Additionally, the COCO dataset lacks the dumpster class, which is a critical component for our project. Therefore, to develop a model capable of recognizing the three classes of objects needed, a custom dataset was created, and the pretrained model was fine-tuned accordingly. Two primary sources were used for the image data: first, images from the industry partner of the project (i.e., Altaroad), and second, the publicly available MOCS (Moving Objects in Construction Sites) dataset Xuehui et al., (2021). The MOCS dataset comprises 41,668 images across 13 categories, including Worker, Tower crane, Hanging hook, Vehicle crane, Roller, Bulldozer, Excavator, Truck, Loader, Pump truck, Concrete transport Mixer, Pile driver, and Other vehicle.

To create a three-class dataset, images from Altaroad were primarily used for the dumpster class, with a few cases where there were humans or vehicles present. A total of 492 top-view images of static dumpsters from Altaroad's clients' sites were carefully selected, covering different times of the day to account for varying luminance. These images were labeled using the open-source platform Labelme. Bounding boxes were drawn around instances of objects of interest, and the labels were saved in the YOLO format.

The human class included workers on site, typically wearing a safety helmet, while the vehicle class included trucks, excavators, lift trucks, loaders, and other related vehicles capable of carrying waste material and dumping it into the dumpster. The dumpster class consisted of top-view images of dumpster bins used for C&DW.

To create a dataset, 500 images were randomly selected from the MOCS dataset and processed before merging with the other set of images. From the 13 original classes of MOCS, the Worker class was selected and renamed as human, along with excavator, truck, and other vehicle, which were used to create the vehicle class in the dataset.

The combined set of images resulted in the final dataset, consisting of 992 images and 2,746 instances of objects across the three classes. The distribution of the number of instances of objects of each class can be seen in Figure 2.

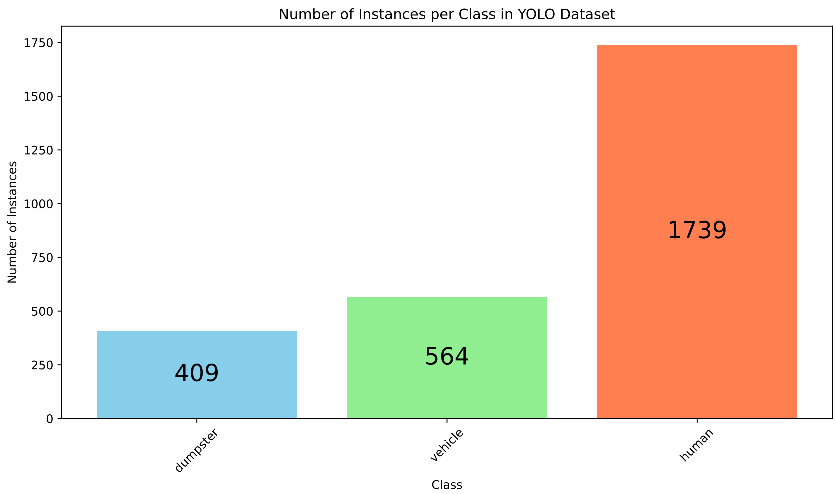


**Figure 2.** Distribution of number of instances of objects in each class of the dataset.

### 3.2.2 Finetuning the Object Detection Model

Larger models with more parameters tend to be more robust in action; however, they have two downsides. First, larger models require more data for fine-tuning, and second, they take more time and computing resources for inference. Therefore, a balance should be maintained between the desired accuracy for inference and the inference time constraint based on the hardware configuration. In this project, the goal was to develop a model capable of running inferences in near real-time on CPUs, so only the nano, small, and medium sizes of the YOLOv8 models were trained and tested.

The hardware setup for the training process included an Nvidia RTX 3090 GPU with 24GB of RAM. The YOLO models have several hyperparameters that can be tweaked for optimal training. The architecture of the model consists of three main parts: the backbone, which is a convolutional neural network based on CSPDarknet and acts as a feature extractor; the neck, which creates feature maps at different scales using a feature pyramid network (FPN) and a path aggregation network (PANet) to enhance information flow; and the head, which is responsible for making predictions using convolutional layers that predict bounding boxes, object scores, and class probabilities (Ju and Cai, 2023).

In our case, since the dataset was relatively small, we only retrained and fine-tuned the part of the YOLO model responsible for classes, while the backbone and neck of the model were frozen. This means that only the head of the model was trained with the data. The built-in data augmentation tool was used during the training process, with different values set for its arguments to create the most accurate model possible.

## 4 Results

The performance of the trained yolov8 nano model was evaluated using precision, recall, F1-score, mean Average Precision (mAP), and a confusion matrix.

The overall F1-score for all classes was 0.73 at a confidence threshold of 0.298, with the mean Average Precision (mAP) at an IoU threshold of 0.5 being 0.740. The model achieved an impressive mAP of 0.991 for dumpsters, indicating nearly perfect detection performance, while the mAP for humans was 0.608 and for vehicles was 0.620, reflecting moderate performance.

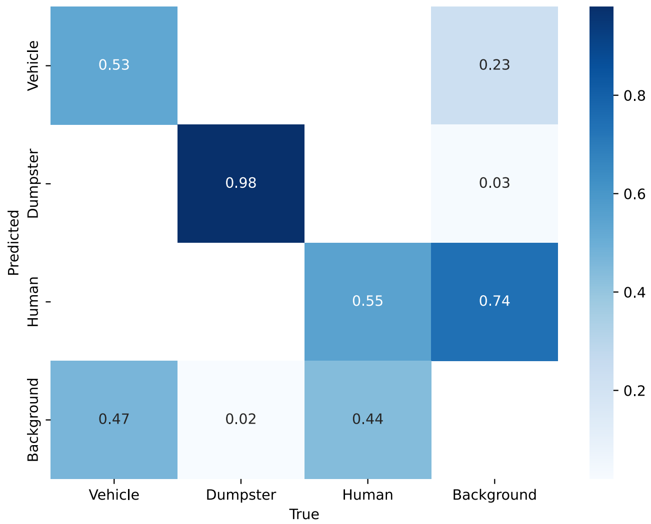


**Figure 3.** Normalized Confusion Matrix for trained model

The normalized confusion matrix, (Figure 3), provides insights into the true positive rates and misclassifications. The model achieved a true positive rate of 0.98 for dumpsters, indicating excellent detection accuracy. However, the true positive rate for humans was 0.55, with significant misclassifications as background, and for vehicles, it was 0.53, also showing notable misclassification as background.

## 5 Discussion

The objective of this research was to develop an automated system for detecting waste disposal events on construction sites. Two primary methods were explored: background subtraction and deep learning-based object detection using the YOLOv8 nano model.

The initial approach using background subtraction demonstrated that it is possible to create a baseline for detecting dumping events. This method relied on segmenting moving objects from a static background to identify potential dumping activities. However, several limitations were identified. The method was highly sensitive to lighting changes and environmental conditions, which often resulted in false positives. Additionally, it required manual positioning of the dumpster in the image for each setup, as the algorithm could not distinguish the dumpster from the background. Due to these limitations, this method was ultimately discarded, highlighting the need for a more robust and automated detection system.

To address the limitations of the background subtraction method, a custom dataset was created for training the YOLOv8 nano model. This dataset included images of dumpsters, humans, and vehicles, labeled to facilitate supervised learning. The creation of this dataset involved collecting images from construction sites and publicly available sources, ensuring a diverse set of scenarios for the model to learn from. However, the dataset's diversity was limited, particularly for the human and vehicle classes, which may have impacted the model's performance.

The YOLOv8 nano model demonstrated promising results, particularly in detecting dumpsters. The overall performance metrics indicated strong precision, recall, and F1-scores for dumpsters, whereas the detection of humans and vehicles showed moderate performance. The confusion matrix analysis further illustrated these findings, showing a high true positive rate for dumpsters but lower rates for humans and vehicles, with significant misclassifications as background (Figure 3). These results suggest that while the model is highly effective at detecting dumpsters, it struggles with identifying humans and vehicles, likely due to the limited diversity in the training dataset for these classes.

The primary challenge of this project was not merely detecting stationary dumpsters but accurately identifying the dynamic events of waste disposal in real-time. By integrating the YOLOv8 nano model with custom programming and scenario-based coding, the system was able to monitor the approach and actions of humans and vehicles towards the dumpsters, effectively identifying dumping events. This combination of advanced object detection and tailored event recognition logic provided a working solution capable of real-time monitoring.

Additionally, the average inference time on a CPU machine with the exported ONNX version of the model was 157ms, demonstrating the model's potential for real-time applications. Future improvements will focus on expanding the dataset diversity to enhance the detection accuracy for humans and vehicles. This includes adding images from different points of view, varying lighting conditions, and diverse environments to better train the model. Additionally, refining the model's architecture and optimizing hyperparameters can further improve its robustness and accuracy. These enhancements aim to create a more reliable and comprehensive solution for real-time construction waste monitoring.

## 6 Conclusion

This research introduced a three-step framework for construction waste monitoring and focused on developing the first module: an automated system for detecting waste disposal events using the YOLOv8 nano model. Despite initial challenges with background subtraction, creating a custom dataset led to significant improvements in detecting dumpsters. However, detecting humans and vehicles showed moderate performance. Future work will expand the dataset's diversity, refine the model's accuracy and robustness, and develop solutions for material recognition and volume estimation, aiming to create a comprehensive and efficient waste monitoring system for construction sites.

## 7 Acknowledgements

This research project was a Mitacs Accelerate International project and was conducted in collaboration with Altaraod Company in France. We extend our sincere gratitude to Mitacs for their financial support, to Altaraod for providing the resources and environment necessary for this work, and finally to École de technologie supérieure (ÉTS) for their invaluable support throughout this research journey.

## References

Bonifazi, G., Capobianco, G., Palmieri, R., Silvia Serranti, S.S., 2019. Hyperspectral imaging applied to the waste recycling sector. Spectrosc. Europe 8. https://doi.org/10.1255/sew.2019.a3

Dong, Z., Chen, J., Lu, W., 2022. Computer vision to recognize construction waste compositions: A novel boundary-aware transformer (BAT) model. Journal of Environmental Management 305, 114405. https://doi.org/10.1016/j.jenvman.2021.114405

Gálvez-Martos, J.-L., Styles, D., Schoenberger, H., Zeschmar-Lahl, B., 2018. Construction and demolition waste best management practice in Europe. Resources, Conservation and Recycling 136, 166–178. https://doi.org/10.1016/j.resconrec.2018.04.016

Hossain, Md.U., Wu, Z., Poon, C.S., 2017. Comparative environmental evaluation of construction waste management through different waste sorting systems in Hong Kong. Waste Management 69, 325–335. https://doi.org/10.1016/j.wasman.2017.07.043

Ju, R.-Y., Cai, W., 2023. Fracture detection in pediatric wrist trauma X-ray images using YOLOv8 algorithm. Scientific Reports 13. https://doi.org/10.1038/s41598-023-47460-7

Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C.L., Dollár, P., 2015. Microsoft COCO: Common Objects in Context. https://doi.org/10.48550/arXiv.1405.0312

Lu, W., Chen, J., 2022. Computer vision for solid waste sorting: A critical review of academic research. Waste Management 142, 29–43. https://doi.org/10.1016/j.wasman.2022.02.009

Menegaki, M., Damigos, D., 2018. A review on current situation and challenges of construction and demolition waste management. Current Opinion in Green and Sustainable Chemistry, Reuse and Recycling / UN SGDs: How can Sustainable Chemistry Contribute? / Green Chemistry in Education 13, 8–15. https://doi.org/10.1016/j.cogsc.2018.02.010

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You Only Look Once: Unified, Real-Time Object Detection. https://doi.org/10.48550/arXiv.1506.02640

Serranti, S., Palmieri, R., Bonifazi, G., 2015. Hyperspectral imaging applied to demolition waste recycling: innovative approach for product quality control. J. Electron. Imaging 24, 043003. https://doi.org/10.1117/1.JEI.24.4.043003

Suzuki, S., Be, K., 1985. Topological structural analysis of digitized binary images by border following. Computer Vision, Graphics, and Image Processing 30, 32–46. https://doi.org/10.1016/0734-189X(85)90016-7

Theophanous, A., n.d. A Vehicle Detection, Counting and Tracking system for Traffic Assessment Using Machine Vision and Aerial Video.

Wang, Z., Li, H., Yang, X., 2020. Vision-based robotic system for on-site construction and demolition waste sorting and recycling. Journal of Building Engineering 32, 101769. https://doi.org/10.1016/j.jobe.2020.101769

Zivkovic, Z., 2004. Improved Adaptive Gaussian Mixture Model for Background Subtraction. Presented at the Proceedings - International Conference on Pattern Recognition, pp. 28-31 Vol.2. https://doi.org/10.1109/ICPR.2004.1333992