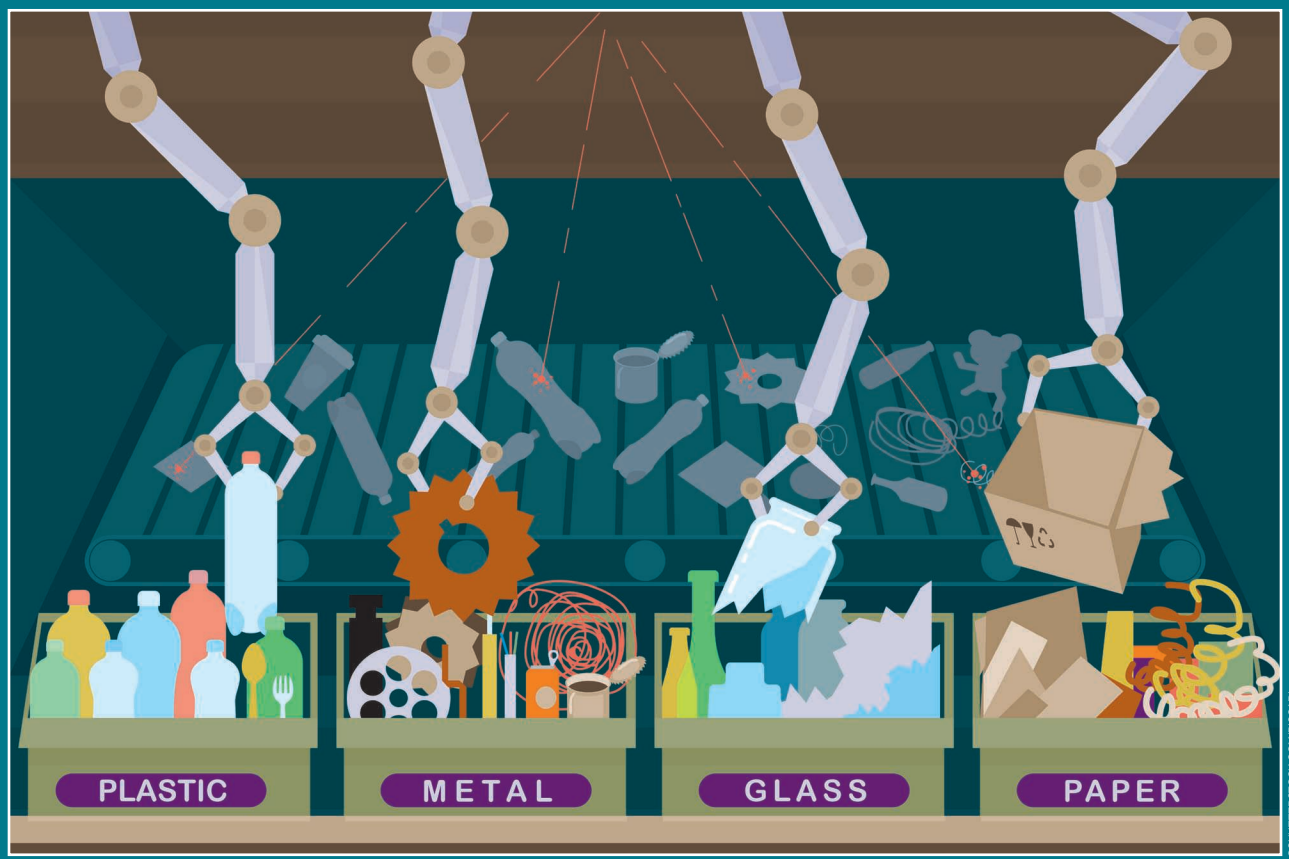
Robotic Waste Sorting Technology

Toward a Vision-Based Categorization System for the Industrial Robotic Separation of Recyclable Waste

By Maria Koskinopoulou, Fredy Raptopoulos, George Papadopoulos, Nikitas Mavrakis, and Michail Maniadakis



The use of robots in waste processing plants can significantly improve the processing of recyclables. Such robots need sophisticated visual and manipulation skills to be able to work in the extremely heterogeneous, complex, and unpredictable waste sorting industrial environment. This article considers the implementation of an autonomous robotic system for the categorization and physical sorting of recyclables according to material types. In particular, it focuses on the development of a low-cost computer vision module based on deep learning technologies to identify and sort items. To facilitate further research endeavors, the data set of recyclable images and a group of image processing scripts for object identification, masking, and synthetic placement against multiple backgrounds are available in an open source GitHub repository (https://github.com/kskmar/ReSort-IT.git). The deep-trained computer vision module is integrated with a robotic system that undertakes the physical separation of recyclables. The composite system is

Digital Object Identifier 10.1109/MRA.2021.3066040
Date of current version: 16 April 2021

1070-9932/21©2021IEEE

deployed in a waste processing plant, where it is successfully assessed in recyclable sorting under difficult and demanding industrial conditions.

The Need for Automation

The current trend toward an environment-friendly "circular economy" changes the way products are made and consumed. The circular economy is based on the assumption that materials will not end up in landfills but will be recovered and reused. Traditionally, the recovery of valuable material from waste has been performed by humans, a solution that suffers from low productivity and increased health risks. During the past two decades, optical sorters have been become popular in industrial material recovery facilities (MRFs). The devices employ a combination of lights and sensors to illuminate and capture images of objects. Image processing reveals the qualitative characteristics of materials, which determine whether objects should be accepted or rejected. Optical sorters use compressed air to move rejected objects to different bins. This technology has difficulty dealing with heavy items, such as semifull containers, which are not rare in waste flows. As a means to overcome the limitations of optical sorters, a new trend involves the use of robotic technology. This approach assumes that there is a low integration cost and that equipment can be easily installed in existing MRFs. Robotic waste sorting systems rely on high-cost imaging technology that exploits recent artificial intelligence (AI) advancements to identify and categorize waste. However, many computer vision solutions are built on proprietary data sets that are not publicly available.

This article supports the development of low-cost, computer vision-based waste categorization modules that can be directly applied in the industry. To this end, we exploit the industrial research setup implemented in the MRF on the island of Crete, Greece (Figure 1), to collect images of different waste types. The images are further processed in a lab to develop a rich and well-documented recyclable waste data set that is used to train an AI-powered module for recyclable detection and categorization. The data set as well as the entire collection of image processing tools is available for public use. The implemented module is integrated with our robotic separator [1], [2] to recover recyclables from urban waste streams. In addition to the integration of the vision-based waste sorting module, the present work discusses the partial advancement of the robot to reduce the time to complete the task. The composite system is tested in demanding industrial operating conditions, with very promising results (an average success rate of 91.8%).

Related Work

Today, there is mounting pressure for more, and more effective, recycling; in the European Union, for example, European Commission Waste Directive 2018/850 stipulates that, by 2035, the amount of municipal waste that is landfilled should be reduced to 10% or less. Industrial waste treatment has, so far, been mostly based on the manual sorting of recyclables for the recovery of valuable, reusable materials. However, this is a laborious and potentially dangerous job that humans should avoid. Recently, many automated systems have been developed that can separate and recover materials, such as metal, paper, glass, and plastic, from waste streams using optical sorters, magnets, eddy currents, and inductive and near-infrared sensing; [3] summarizes advances in the mechanical processes, sensors, and actuators that are employed. Robotic technology is coming to support (and even replace) existing installations by offering a more effective and autonomous alternative to the established procedure [4], [5]. Currently, the most well-known commercial robots used in the waste management industry come from Sadako [6], SamurAI [7], AMP Robotics [8], and ZenRobotics (which also deals with the management of construction waste) [9]. The majority of the existing systems capitalize on the agility of

Robots need sophisticated visual and manipulation skills to be able to work in the extremely heterogeneous, complex, and unpredictable waste sorting industrial environment.

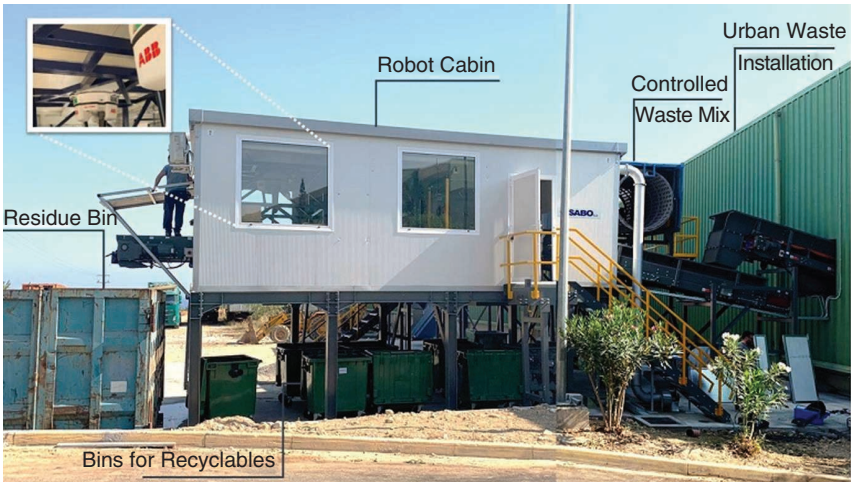


Figure 1. A depiction of the industrial research installation at the Crete MRF. Starting from the right, two feeders place waste on a conveyor belt. The feeders are located above the belt to provide a controlled waste mixture or inside the processing installation to feed real urban waste. The belt carries waste to the robot cabin for sorting. Under the cabin, there are separate bins for each material. The residue of the sorting process ends up in a dedicated bin.

delta robots to rapidly transfer recyclables from a conveyor belt to a bin.

Besides commercial systems, recent research endeavors aim to improve automated recyclable recovery technology. For example, Zhihong et al. [10] used an industrial KUKA manipulator with a two-finger system to sort waste according to the vision-based classification of moving objects. Similarly, we recently developed a robotic system that accomplishes one sort per second, utilizing delta robot technology [2]. Bircanoğlu et al. [11] assessed well-known deep convolutional neural network (CNN) architectures for the classification of different types of recyclables. Additionally, the authors of [11] proposed a deep learning architecture for recyclables classification, namely RecycleNet, which has been shown to be very effective in categorizing images of single objects. In contrast to the preceding, the current work focuses on much more complex multiobject images, solving a three-fold problem related to object identification (bounding box specification), localization (masking), and material type attribution (classification). This triplet makes our solution applicable in industrial conditions where several potentially overlapping recyclables shown in the very same image need to be identified, localized, and classified.

Robotic technology is coming to support (and even replace) existing installations by offering a more effective and autonomous alternative.

Industrial Research Setup

The development of a robotic system that can be applied in industrial environments presupposes extensive assessments in realistic and demanding conditions. To this end, with the help of waste management experts, a research setup was implemented at Crete's MRF (Figure 1) to facilitate data collection, experimentation, and system assessment. Key components deployed in the robot cabin (see Figure 1) are described in the following.

- *Conveyor belt*: Typically, industrial waste processing units transfer recyclables on conveyor belts. We developed a replica of such a setup through the installation of a 22.5-m long, 1-m wide (usable width: 0.8 m) belt with a speed of up to 0.15 m/s and an optical encoder system so that the speed could be monitored.
- *Waste feeder*: A dual feeder shed waste at a controlled rate onto the belt. The material came from a waste processing installation that transferred uncontrolled urban waste or a from a newly developed stream located above the conveyor to feed the system a controlled waste mix for experimentation purposes.
- *Camera*: A stereo, full-high-definition ZED camera enabled the automated categorization of recyclable materials (see the "Vision-Based Material Categorization" section). The camera was placed 148 cm in front of the robot (in the direction in which the recyclables were moving), at a height of 75 cm above the conveyor belt, looking downward. To ensure constant light conditions during data acquisition and system operation, the camera was placed inside a box with LED equipment. The camera's field of view covered the entire width of the conveyor belt, providing information about the shape and color of the transported waste.
- *Robot*: An ABB IRB360 delta robot was installed above the conveyor belt to enable the picking and transfer of waste to bins. The robot consisted of three high-torque servomotors mounted on a rigid frame, each linked to a different arm. The three arms were connected to a central platform, driving it to move very fast and accurately in 3D space. The robot had a payload of 6 kg, which was appropriate for repetitive and rapidly completed applications.
- *Vacuum gripper*: To enable automated picking and placing, we attached a vacuum gripping module to the end effector of the robot; it consisted of a vacuum blower that provided high-volume suction to pick up and hold selected materials (see the "Recyclable Object Manipulation" section). The use of vacuum technology provides a robust and low-cost solution for material transfer [1].

Material Detection and Manipulation

This article presents an integrated robotic system for recyclable sorting that is composed of two main parts: a robotic manipulator for the physical separation of waste to different bins, depending on the material type, and a vision-based material detection and categorization module.

Picking and Placing Moving Recyclables

In the recycling industry, waste is typically preprocessed to facilitate material recovery. In our setup, recyclables are filtered through a trommel (a rotating punch plate) to remove small objects. At the end of the trommel, recyclables fall onto a conveyor belt to be transported to the robotic sorting installation. The conveyor belt moves at a constant velocity, facilitating the real-time estimation of the location of identified objects (see the "Vision-Based Material Categorization" section) and thus the planning of the robot's activities.

Any object added to the recognized recyclables list is a potential target for the robot. Let $P_0 = (x_0, y_0, z_0)$ be the robot's starting position and $P_t = (x_t, y_t, z_t)$ be the target's position (the reference frame in Figure 2). The conveyor moves along the x-axis at a constant velocity of $v = 0.2$ m/s. Then, the location of the moving target is computed by

$$\begin{cases} x_{\text{pick}} = x_t + (t_{\text{all}} \times v) \\ y_{\text{pick}} = y_t \\ z_{\text{pick}} = z_t \\ t_{\text{all}} \ \ = t_H + t_c \end{cases}, \tag{1}$$

where t_{all} is the total picking time, t_H denotes the duration of the robot's horizontal displacements, and t_c is the robot's idle time and vertical movement time. The robot's acceleration a_r is set to 150 m/s^2, and $C_v = 2$ is a constant parameter that describes the motion profile. Hence,

$$\begin{cases} D &= \sqrt{(x_{\text{pick}} - x_0)^2 + (y_{\text{pick}} - y_0)^2} \\ x_{\text{pick}} &= x_t + (t_{\text{all}} \times v) \\ t_{\text{all}} &= t_H + t_c \\ t_H &= (C_v D/a_r)^{1/2} \end{cases}. \tag{2}$$

Then, by simplifying the latter equation and replacing the differences $\Delta x = x_t - x_0$ and $\Delta y = y_{\text{pick}} - y_0$, (2) becomes

$$D^2 = (v \times (C_v D/a_r)^{1/2} + v \times t_c + \Delta x)^2 + \Delta y^2. \tag{3}$$

Equation (3) is not trivial to be solved. However, we apply Newton's iterative method for solving differential equations, as proposed in [12], which provides the target position:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}. \tag{4}$$

By this process, the picking position is resolved, and the robot can generate the trajectory.

Recyclable Object Manipulation

The majority of robotic systems for recyclable sorting rely on a pick-and-place (PnP) process to select objects and physically transfer them to the appropriate material bin. Capitalizing on delta robots, we developed a composite system that performs fast recyclable sorting by adopting the PnP approach [1], [2]. Given that recyclables do not require fine manipulations, the gripping of objects is often implemented with the use of vacuum technology. Generally speaking, there are two main technologies to generate a vacuum. Venturi generators use compressed air flowing through a conical orifice to develop a pressure difference. Their main advantages include a short response time that enables fast picking and a powerful gripping force once the suction cup has sufficiently sealed on an object. The use of blowers provides an alternative. Blowers spin synchronized rotors in a chamber to generate a vacuum that is relatively weak but that has a high volume (typically more than 500 m^3/h). The problem with using blowers for recyclable picking lies in the equipment's slow response time, which may significantly decelerate waste processing. For this reason, Venturi generators were used in our previous work [1], [2].

Testing the composite system in laboratory conditions, where recyclables are less deformed, yielded excellent recyclable manipulation performance. However, after transferring the system to, and assessing it in, industrial conditions, it was occasionally impossible for the suction cup to form a seal with objects, resulting in failures to grab items. To compensate for this, the present work incorporates a blower. The high volume of pumped air increases the system's ability to pick up significantly deformed objects when there is no perfect suction cup seal. This comes with the cost of slowing the processing speed because the blower valve used to activate/deactivate the vacuum at the end effector has a sluggish response, on the order of 500 ms/cycle. Therefore, we examined methods to gain speed.

The development of a robotic system that can be applied in industrial environments presupposes extensive assessments in realistic and demanding conditions.

When picking, the delay was reduced by sending the vacuum activation command at a properly timed moment while the robot was moving toward the goal. Moreover, to decrease the material disposal delay, we developed a solution that forced compressed air toward an item to detach it from the suction cup. Specifically, a custom T-shaped tube enabled fast switching between the blower vacuum (negative pressure) and the compressed air (positive pressure), right on the suction cap (see Figure 2). The joint activation/deactivation of binary valves controlling the two air streams (vacuum versus compressed air) significantly reduced the time needed for item disposal. When the robot's end effector reached the bin, the two valves turned off the vacuum and activated the compressed air, thus making items detach from the suction cup in much less time.

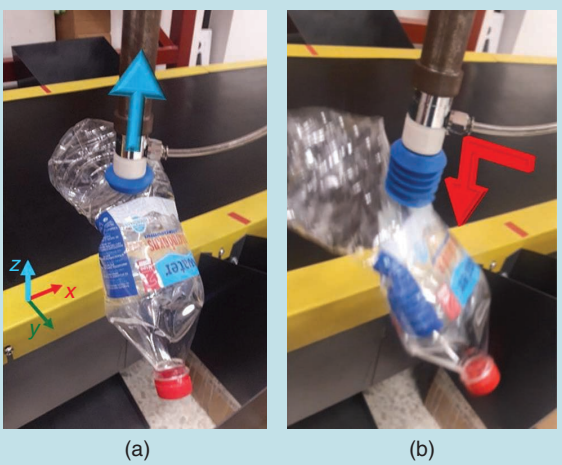


Figure 2. Two types of air flow. (a) Vacuum suction (blue arrow). (b) Compressed air (red arrow) used for disposal.

Vision-Based Material Categorization

Industrial consumers of recovered materials maintain specifications for purity and quality, as recyclables are the main ingredients for many products. Accordingly, any autonomous recovery system needs to accurately categorize recyclables based on their material type. One way to achieve this is to use hyperspectral cameras, although this significantly increases the cost of composite systems. Alternatively, one may exploit recent developments in deep learning to come up with an efficient and much cheaper solution. CNNs trained with deep learning algorithms have been widely applied in demanding computer vision applications [13]. Interestingly, CNNs have also been used to perform waste classification [14], although they were applied to a relatively small data set with single-item images and without any type of occlusion. In contrast, the present work targets "instance segmentation" to accomplish multiple-object identification and labeling. To this end, the well-known Mask Regional CNN (R-CNN) [15] open source network was employed since it has been particularly successful at similar tasks. Mask R-CNN provides a scalable means for categorizing recyclables into a high number of classes. However, this cannot be done incrementally. Any time a class is added, a full retraining of the model is assumed, according to the specification of the new data set.

Industrial consumers of recovered materials maintain specifications for purity and quality, as recyclables are the main ingredients for many products.

Synthetic Data Generation Process

Implementing a rich data set that consists of many recyclable waste images of different types is the first important step in creating the categorization module. To the best of our knowledge, there are only two open source data sets, namely, TrashNet [16] and Taco [17], that are available for public use. Both cover waste classification for outdoor cases rather than for demanding industrial setups, where objects transferred on conveyor belts can be strongly deformed, dirty, and piled on top of one another. The present work's approach to develop the data set differs from the techniques in the works mentioned previously. Typical image data collection methods assume that photos of objects of interest will be taken against a variety of backgrounds and manually labeled using annotation tools [13], [14]. This is a complicated and time-consuming approach that can hardly support the collection of very large data sets. Instead, this article adopts an inverse strategy that aims to automate the annotation process and thus develop an arbitrary amount of data that may address different levels of problem complexity. The steps we took to develop a rich waste data set are outlined in the following.

The resulting mask was assigned as an image annotation to clearly indicate the region of interest; it was also used for Mask R-CNN training [Figure 3(a)]. In the second step, we randomly applied three geometric transformations (translation, rotation, and scaling) to the image and the mask. This resulted in a large data set of single-object images against a black (i.e., a conveyor belt) background. Finally, pairs of objects from the second step were randomly selected (with their masks) and arbitrarily placed over new images with colorful backgrounds to develop complex problem instances. In the case of recyclable object data collection, this step provided a means to examine cases with multiple, possibly overlapping, objects across a range of diverse backgrounds [Figure 3(b)].

Figure 3. (a) The basic data set, including representations of each class of studied material: aluminum (first row), paper and cardboard (second row), PET bottles (third row), and nylon (fourth row). (b) The augmented data set.

Data Acquisition

We followed the semiautomated procedure summarized earlier to develop the data set for training Mask R-CNN. We started by collecting 400 images for each of the material types considered in the study, namely, aluminum [Figure 3(a), first row], paper and cardboard [Figure 3(a), second row], polyethylene terephthalate (PET) bottles [Figure 3(a), third row], and nylon [Figure 3(a), fourth row]. Then, we followed a manual procedure to collect red–green–blue images. All images had the same size, i.e., 800 × 800 pixels, and depicted a single recyclable against a black background. It is noted that, to have a representative data set of recyclables images, we also considered deformed and dirty materials. These images were processed by following step one (mentioned previously) to identify masks describing regions of the identified objects. We called this the Basic data set. Then, we applied step two (from the preceding) to gather 2,000 images of each material type. We called this the Synthetic Single data set. Finally, following the third processing step, we developed a data set that artificially described complex cases of multiple and overlapping objects. This was the Synthetic Complex data set. In particular, we developed 16,000 images for the training set and 5,600 for the validation set. Figure 4 includes four random images from the data set that have been successfully created and annotated following the data augmentation process.

Experimental Results

The robotic module and the material categorization module are integrated into a composite system that is deployed and validated in the industrial research setup summarized in the "Industrial Research Setup" section. The two modules run in full parallel, with the visual detection module adding object locations to the list of potential targets managed by the robot.

Validation of Recyclable Robotic Manipulation

To assess the success of the two alternative vacuum gripping systems, i.e., the Venturi generator and the blower, we examined their ability to pick up 80 representative recyclables (20 for each class) that consisted of aluminum cans in flat, crumpled, and cylindrical shapes; paper and nylon in flat, crumpled, and ball shapes; and PET bottles in flat, arbitrarily crumpled, and cylindrical shapes, some of which contained liquid. Indicative object shapes appear in Figure 3(a). The experimental procedure ran in a fully automated mode through five cycles using the composite set of 80 representative recyclables (400 picks in total). We used the Mask R-CNN recyclable identification module (see the following) to identify the location of objects on the moving conveyor belt. Table 1 summarizes the results, documenting the ability of each vacuum system to pick up and transfer items (successful and unsuccessful trials, respectively). As shown in the table, in most cases, the blower outperformed the Venturi

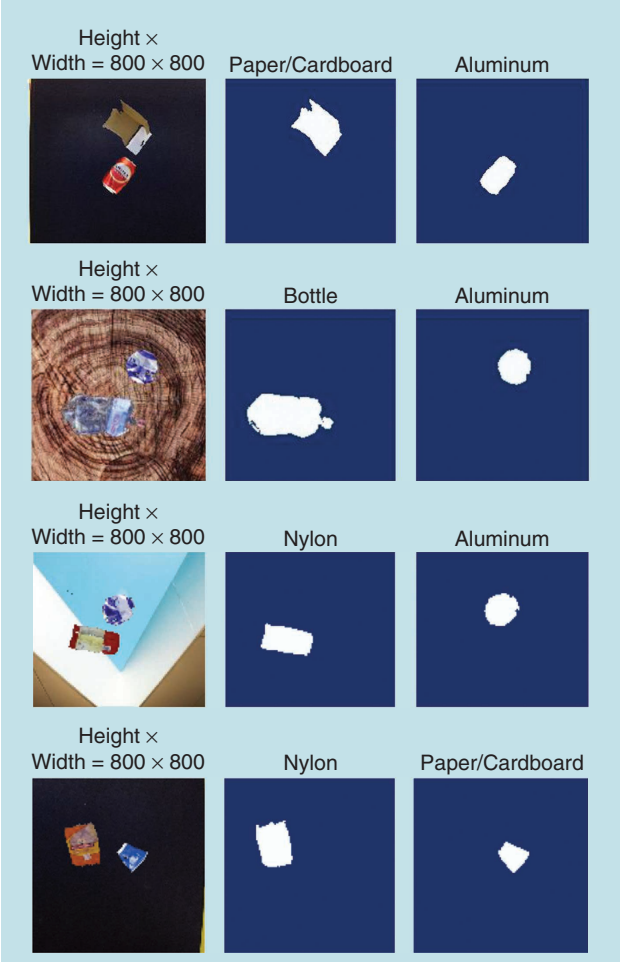


Figure 4. Four samples from the training data set presented with annotated masks. The first column contains original images. The second and the third columns show each material mask and its class identification as a subtitle.

Table 1. The PnP performance evaluation (%).

Approach		Aluminum	Paper/Cardboard	Bottle	Nylon	Average
Venturi	Pick success	68	85	81	98	83
	Transfer fail	0	1	3	0	1
Pump	Pick success	**82**	**96**	**84**	**99**	**90.2**
	Transfer fail	0	0	4	0	1

generator. This was particularly true for lightweight recyclables. However, when we considered heavy items (e.g., 2-kg PET bottles full of water), the blower was unable to grab the objects. In contrast, the Venturi solution was much more successful because of its ability to form an effective seal. The two systems produced similar results in the case of 1-kg PET bottles, but the blower outperformed the Venturi when dealing with 0.5-kg and empty bottles, which typically have unstructured surfaces. After examining the results and considering the fact that the vast majority of recyclables are lightweight, we selected the blower.

Validation of Visual Categorization

Training Process

The development of a vision-based module capable of categorizing recyclables into different material types was accomplished using Mask R-CNN [15], which has been successfully employed to tackle a wide range of object identification and categorization tasks. Mask R-CNN is a deep CNN that simultaneously predicts recyclable object bounding boxes, masks, and material types. We used a public Mask R-CNN implementation [18] that was trained on the data sets in the "Vision-Based Material Categorization" section, using 30% of the images for model learning validation and the remaining 70% for model training. To increase the image processing frame rate, we incorporated ResNet-50 for feature extraction across entire photos. Moreover, a region proposal network (RPN) was employed to scan images in a sliding window manner to identify areas that contained objects. The RPN employed a set of boxes, called *anchors*, with predefined image locations and scales (there were five scales and three aspect ratios in our implementation) to figure out the size and location of an object on the feature map. Specifying the correct anchor span is critical for obtaining successful classification results. To estimate recyclable objects' bounding boxes, masks, and material types, dedicated subnetworks known as *heads* work on identified regions of interest to shape the final output. After extensive experimentation with the validation set, we identified the Mask R-CNN training parameters that worked effectively. They are summarized in Table 2.

Table 2. The training parameter configuration.

Parameter	Value
BACKBONE	ResNet-50
BATCH_SIZE	1
DETECTION_MIN_CONFIDENCE	0.85
FPN_CLASSIF_FC_LAYERS_SIZE	1,024
RPN_ANCHOR_RATIOS	[0.5, 1, 2]
RPN_ANCHOR_SCALES	(32, 64, 128, 256, 512)
RPN_ANCHOR_STRIDE	1
RPN_BBOX_STD_DEV	[0.1 0.1 0.2 0.2]
RPN_NMS_THRESHOLD	0.7
RPN_TRAIN_ANCHORS_PER_IMAGE	256

Before using Mask R-CNN for recyclable identification and classification, a learning process was applied to improve the applicability of the deep neural network to the problem. To bootstrap the learning process, a network pretrained on the Common Objects in Context data set was used. This approach, introduced in [19], speeds up learning and ensures the minimum quality of the results in a reasonable training time. Following this, only the Mask R-CNN heads are typically trained for a problem, keeping the general object feature extraction mechanism unchanged. To adapt Mask R-CNN to the classification task, we developed and tested four customized implementations: Net 1, Net 2, Net 3, and Net 4. They were trained on a high-performance computing infrastructure with TensorFlow GPU support.

Training Analysis

Mask R-CNN training is based on a complex loss function that is calculated as the weighted sum of different partial losses at every training state. The partial losses considered in the current implementation are described in the following.

- *mrcnn_bbox_loss* corresponds to the success at localizing the bounding boxes of objects that belong to a given class. This loss metric is increased when the object categorization is correct but the bounding box localization is not precise.
- *mrcnn_class_loss* describes the loss due to the improper categorization of an object identified by the RPN. This provides an indication of how well Mask R-CNN recognizes each class of an identified object. This metric is increased when an object is detected in an image but misclassified.
- *mrcnn_mask_loss* is the metric that summarizes the success at implementing masks across a processed image that correspond to the pixel area covered by the identified objects. The training data set includes masks for all objects of interest in input images, thus providing the ground truth for the evaluation of masks predicted by Mask R-CNN.
- *rpn_bbox_loss* corresponds to the localization accuracy of the RPN, in other words, how well the RPN identifies individual objects. High values indicate miscalculations when an object is detected but the bounding box needs to be corrected.
- *rpn_class_loss* is an RPN performance metric that is increased when an object is undetected at the final output and decreased when the RPN successfully identifies an item.

Performance Analysis

To assess the performance of each trained network, we used a typical evaluation procedure that considered three metrics,

namely, average precision (AP), average recall (AR), and the F1 score (F1), which are calculated as follows:

$$AP = \frac{tp}{tp + fp}, \tag{5}$$

$$AR = \frac{tp}{tp + fn}, \tag{6}$$

$$F1 = 2 * \frac{AP \cdot AR}{AP + AR}. \tag{7}$$

In these equations *tp*, *fp*, and *fn,* respectively, denote the true positive, false positive, and false negative identification of recyclables. Therefore, AP measures the percentage of correct positive predictions among the total positive predictions of a class, AR denotes the percentage of correct positive predictions among the actual number of the class items, and F1 is the harmonic mean of the precision and recall. Additionally, we calculate the mean values of AP (MAP) and AR (MAR) to provide a picture of the overall Mask R-CNN performance.

Implementation of the Mask R-CNN Model

Besides tuning the training procedure parameters, there are conditions that may affect the quality of the obtained results. These include the data set size, the number of the learning steps per epoch, and the use of additional data augmentation strategies to improve generalization. Accordingly, several Mask R-CNN training variations were examined for their training and validation results (Table 3). The inference performance of the obtained models on 200 new test images captured from the conveyor belt is described in Table 4. The different versions of the examined models are summarized in the following.

- *Net 1*: We began with the relatively small Basic data set, which consisted of single-object, black-background images (see the "Vision-Based Material Categorization" section). The training process ran with 50 training steps per epoch, for a total of 200 epochs (this setup was also followed in the next two cases). Although the network was successfully trained (Table 3), an assessment against real industrial images revealed many limitations that rendered this solution useless in practice. This is clearly indicated in Table 4, which highlights the recall, precision, and F1 performance of the network for each material and in total. The limited success of this version provided the driving force for developing artificial data augmentation algorithms.
- *Net 2*: The next attempt considered the much larger Synthetic Single data set for training Mask R-CNN. The solution yielded slightly better results for identifying individual items, as demonstrated in Table 4. However, the average F1 score remained low, proving poor overall performance. This was mainly because the training data set presented recyclables against a black (conveyor belt) background. Thus, when this implicit hypothesis was violated (i.e., due to object occlusion), the network performance dropped dramatically.
- *Net 3*: A major improvement to the data set concerned the synthetic placement of objects against colorful backgrounds that could also contain other objects. The Synthetic Complex data set was used to train a new Mask R-CNN (the loss metrics are summarized in the third row of Table 3), which provided the first vision-based recyclable identification and categorization solution that could be potentially applied in the real world. As shown in Table 4, the network's success rate in demanding industrial conditions significantly increased over that of Net 1 and Net 2.
- *Net 4*: To further enhance the Mask R-CNN performance, we made changes in two directions that improved generalization and robustness. First, random affine transformations

Table 3. The training and validation losses (minimum values).

	Network	Mask R-CNN Bounding Box	Mask R-CNN Class	Mask R-CNN Mask	RPN Bounding Box	RPN Class
Training	Net 1	0.22	0.21	0.21	0.41	0.04
	Net 2	0.21	0.15	0.21	0.4	0.04
	Net 3	0.2	0.12	0.2	0.38	0.04
	Net 4	0.1	0.09	0.18	0.22	0.02
Validation	Net 1	0.18	0.18	0.21	0.88	0.04
	Net 2	0.17	0.14	0.22	0.73	0.04
	Net 3	0.17	0.12	0.18	0.81	0.04
	Net 4	0.1	0.09	0.11	0.21	0.02

Table 4. The Mask R-CNN performance (%) on material classification with industrial images.

	Aluminum			Paper/Cardboard			Bottle			Nylon			Total Average		
Network	AP	AR	F1	AP	AR	F1	AP	AR	F1	AP	AR	F1	MAP	MAR	F1
Net 1	69.5	69.5	69.5	46.3	55.5	50.5	57.5	63.2	60.2	35	41	37.8	52.1	57.3	54.5
Net 2	72.9	76.5	74.7	63.8	71.5	67.5	75.7	76.5	76.1	45	49.5	47.1	64.4	68.5	66.3
Net 3	81.5	81.5	81.5	65.5	72	68.6	74.8	81.5	78	59.5	65.5	62.4	70.3	75.1	72.6
Net 4	**94.1**	**95**	**94.5**	**81.4**	**89.5**	**85.2**	**91.3**	**94**	**92.6**	**81.9**	**88.5**	**85.1**	**87.2**	**91.8**	**89.4**

were performed on the Synthetic Complex data set, i.e., translation, rotation, and scale in the ranges (−0.2, 0.2), (−45, 45), and (0.8, 1.2), respectively. Practically, this corresponds to random augmentations of the data set to improve generalization. Additionally, the training process was extended by using 100 learning steps in each epoch. The individual loss metrics through the training process are summarized in the bottom row of Table 3. Interestingly, the new solution outperformed Net 3 when tested on industrial images, as seen by comparing the bottom two rows of Table 4.

Figure 5 presents three indicative snapshots from the testing images (first column) and the corresponding outcome for each testing architecture: Net 1 [Figure 5(b)], Net 2 [Figure 5(c)], Net 3 [Figure 5(d)], and Net 4 [Figure 5(e)].

Further Assessment

Having established a good basis for the visual identification and categorization of recyclables, it is important to elaborate on the evaluation of the system. To this end, we considered an enhanced testing data set that consisted of 1,000 random real-flow images depicting recyclables (potentially with overlapping parts). We systematically examined the effect of affine transformations that are shown to affect Mask R-CNN performance [20]. In particular, we employed the well-known image augmentation tool imgaug (https://github.com/aleju/imgaug) to examine the isolated effect of the different transformations on the training process and reveal the ones that provided more benefits for a given application domain. The results are reviewed in Table 5. Clearly, the performance of the networks with a single transformation operator was poorer compared to the networks trained with combinations. Moreover, in most cases where a single transformation was used, the MAP (the precision metric) was significantly lower than the MAR (the recall metric), due to increased false negative categorizations.

Interestingly, in contrast to translation, the rotation and scaling operators had a more positive effect on network performance, reaching MAR scores of 80.1% and 84.2%, respectively. At the same time, Gaussian blur did not enhance the model training. This might be because the data set used for training consisted of blurred images, leading the extra blur to

Figure 5. Material classification using Mask R-CNN. A dotted bounding box and a color-filled mask are assigned to every predicted material. The name of the class and the prediction confidence are also depicted for each classified material. Colors are randomly selected. (a) Test input image. (b) Net 1 predictions. (c) Net 2 predictions. (d) Net 3 predictions. (e) Net 4 predictions.

destroy information. The performance of Net 4 across the entire set of multiple-object images was still higher than all the investigated solutions. This is an indication that the combination of rotation, translation, and scaling contributes the most to improving Mask R-CNN performance.

From the Lab to the Productive World

Following the evaluation of the individual modules, a composite system was implemented that consisted of the delta robot equipped with the vacuum gripper and the Net 4-based computer vision module. The system was deployed in Crete's industrial MRF (see Figure 1) to physically sort recyclables into dedicated bins. The performance of the robotic sorter is demonstrated at https://bit.ly/2INJq6W. Besides the fact that the data set was not sufficiently rich in terms of object variability (i.e., a finite number of images is shuffled to produce a large data set), the system correctly identified and categorized the majority of the considered objects. To further reduce the number of false positives, the robot treated only objects that were recognized at a very high level of confidence (i.e., more than 98%). Moreover, the pump-based vacuum was very effective at grabbing and transporting lightweight objects. Overall, the robot could effectively complement the manual sorting of recyclables, thus having a positive impact on the productivity of the MRF. The fast, precise, and tireless operation of the robot could reduce the amount of waste that workers had to treat, thus facilitating the processing of the large volume of material that arrives daily at the plant.

Discussion and Future Work

The industrial research setup in the Crete MRF crucially supported experimentation toward the development of a robotic waste sorter that is applicable in the real world. However, there are points that have to be clarified to help the reader better appreciate the results of the current work. As expected, despite the increased ability of the computer vision module to detect and sort recyclables in complex synthetic images, when applied in the real world, the system's success declined, due mostly to uncontrolled dirt on the recyclables. It is noted that the experiments took place during summer, which means there was little mud on the waste to effect detection and sorting. Interestingly, experts pointed out that waste exhibits seasonal characteristics that can affect management strategies. For example, in the summer, waste flows include more PET bottles that contain liquid, which cannot be treated by optical sorters. This highlights the benefit that the robotic system could provide to industrial waste treatment.

At the same time, as observed in Table 3, there are types of recyclables for which the Mask R-CNN-based solutions achieved high categorization scores. In particular, the identification of aluminum cans and PET bottles was more successful in comparison to cartons and nylon. This is because the shape and color variety of aluminum cans and PET bottles are more constant than those of other classes. Interestingly, in many images, the distinction between cartons and nylon is so difficult that even the human eye could be deceived.

Due to the fact that recovered PET bottles have high value in the recyclables market, the waste management experts asked for the development of a dedicated system with binary PET/no-PET functionality. This is the goal that drives our midterm future work. Given that, on Crete, 18 types account for more than 85% of the PET bottles that go to the plant, the presented results are already very encouraging. In a different direction, there is already interest in installing a pair of robots that will work together to manage a larger volume of waste by taking on separate or complementary roles, depending on the needs of the plant. This is a topic that we will study in the future.

There is already interest in installing a pair of robots that will work together to manage a larger volume of waste by taking on separate or complementary roles.

Conclusions

Current solid waste management practices are no longer sustainable. With the exploitation of state-of-the art computer vision and robotics technology, the goal of finding innovative ways to attain sustainable waste management becomes more achievable and realistic. This work presented an intelligent autonomous system to support the recovery of recyclable materials. The system was gradually integrated into the operation of an MRF to support the treatment of waste processed on a daily basis. In comparison to previous works, this article accomplished the following:

- It introduced a blower-based vacuum system to improve the ability of robotic systems to manipulate recyclables.

Table 5. The resulting performance (%).

		Real Flow		
Augmentation	**Range**	**MAP**	**MAR**	**F1**
Translation (x, y)	(−0.2, 0.2)	68.5	76.7	72.2
Translation (x, y)	(−0.5, 0.5)	65.4	72.9	68.8
Rotation	(−20, 20)	70.3	78.8	74.2
Rotation	(−40, 40)	71.8	80.1	75.6
Scale	(0.9, 1.1)	72.4	79.4	75.7
Scale	(0.7, 1.3)	75.3	84.2	79.5
Gaussian_Blur	Sigma = 2	56.2	57.6	56.9
Gaussian_Blur	Sigma = 3	48.4	42.2	45.1
Translation + Rotation + Scale + Blur	(−0.2, 0.2) + (−30, 30) (−0.8, 1.2) + 2	71.3	73.4	72.3
Translation + Rotation + Scale	Same as Net 4	**87.4**	**90.1**	**88.7**

- It provided a new data set of recyclable images and a group of processing tools to facilitate deep learning research into recyclable categorization.
- It developed a composite autonomous system that successfully identified, localized, and categorized recyclables in a demanding industrial environment.

The technology proposed here could significantly impact future waste treatment plants, which are envisioned to be highly automated facilities where no humans will be in direct contact with waste and where almost all recyclables will be recovered.

Acknowledgments

This research was cofinanced by European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship, and Innovation, under the call RESEARCH–CREATE–INNOVATE (project name: ANASA; code:T1EDK-03110), MIS 5031867.

References

[1] E. Papadakis, F. Raptopoulos, M. Koskinopoulou, and M. Maniadakis, "On the use of vacuum technology for applied robotic systems," in *Proc. ICMRE 2020—IEEE 6th Int. Conf. Mechatron. Robot. Eng.*, 2020, pp. 73–77.

[2] F. Raptopoulos, M. Koskinopoulou, and M. Maniadakis, "Robotic pick-and-toss facilitates urban waste sorting *," in *Proc. IEEE 16th Int. Conf. Automat. Sci. Eng. (CASE)*, 2020, pp. 1149–1154.

[3] R. Sarc, A. Curtis, L. Kandlbauer, K. Khodier, K. Lorber, and R. Pomberger, "Digitalisation and intelligent robotics in value chain of circular economy oriented waste management—A review," *Waste Manage.*, vol. 95, pp. 476–492, July 15, 2019. doi: 10.1016/j.wasman.2019.06.035.

[4] E. Mokled, G. Chartouni, C. Kassis, and R. Rizk, "Parallel robot integration synchronization a waste sorting system," in *Mechanism, Machine, Robotics and Mechatronics Sciences* (Mechanisms and Machine Science), R. Rizk and M. Awad, Eds. Cham: Springer-Verlag, May 2019, pp. 171–187.

[5] Z. Zhang, H. Wang, H. Song, S. Zhang, and J. Zhang, "Industrial robot sorting system for municipal solid waste," in *Intelligent Robotics and Applications*, H. Yu, J. Liu, L. Liu, Z. Ju, Y. Liu, and D. Zhou, Eds. Cham: Springer International Publishing, 2019, pp. 342–353.

[6] Sadako. Accessed Oct. 2020. [Online]. Available: http://www.sadako.es/max-ai/

[7] "SAMURAI robotic sorting system." SAMURAI, Canada. Accessed Sept. 2020. [Online]. Available: https://www.machinexrecycling.com/products/samurai-sorting-robot/

[8] AMP Robotics. Accessed Oct. 2020. [Online]. Available: https://www.amprobotics.com/

[9] Zen Robotics. Accessed Oct. 2020. [Online]. Available: https://zenrobotics.com/

[10] C. Zhihong, Z. Hebin, W. Yanbo, L. Binyan, and L. Yu, "A vision-based robotic grasping system using deep learning for garbage sorting," in *Proc. Chinese Control Conf. (CCC)*, 2017, pp. 11,223–11,226.

[11] C. Bircanoğlu, M. Atay, F. Beşer, Ö.Genç, and M. A. Kizrak, "Recyclenet: Intelligent waste sorting using deep neural networks," in *Proc. Innov. Intel. Syst. Appl. (INISTA)*, 2018, pp. 1–7.

[12] G. Chen, L. Zhai, L. Li, and J. Shi, "Trajectory planning of delta robot for dynamic tracking, pick and placement," *Adv. Mater. Res.*, vol. 680, pp. 473–478, Apr. 2013. doi: 10.4028/www.scientific.net/AMR.680.473.

[13] N. Sharma, V. Jain, and A. Mishra, "An analysis of convolutional neural networks for image classification," *Procedia Comput. Sci.*, vol. 132, pp. 377–384, 2018. doi: 10.1016/j.procs.2018.05.198.

[14] R. A. Aral, R. Keskin, M. Kaya, and M. Haciömeroğlu, "Classification of trashnet dataset based on deep learning models," in *Proc. IEEE Intern. Conf. Big Data*, 2018, pp. 2058–2062.

[15] F. Massa and R. Girshick, "MaskRCNN-benchmark: Fast, modular reference implementation of instance segmentation and object detection algorithms in PyTorch," GitHub, San Francisco, 2018. [Online]. Available: https://github.com/facebookresearch/maskrcnn-benchmark

[16] G. Thung and M. Yang, "Classification of trash for recyclability status," *CS229 Project Rep.*, Computer Science, Corpus ID: 27517432, cs229.stanford.edu, 2016.

[17] P. F. Proença and P. Simões, "TACO: Trash annotations in context for litter detection," 2020, arXiv:2003.06975.

[18] W. Abdulla, "Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow," GitHub, San Francisco, 2017. https://github.com/matterport/Mask_RCNN

[19] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 2980–2988. doi: 10.1109/ICCV.2017.322.

[20] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," in *Proc. Int. Conf. Learn. Represent.*, 2018. [Online]. Available: https://openreview.net/forum?id=r1Ddp1-Rb

Maria Koskinopoulou, Institute of Computer Science, Foundation for Research and Technology–Hellas, Heraklion, 711 10, Greece. Email: mkosk@ics.forth.gr.

Fredy Raptopoulos, Institute of Computer Science, Foundation for Research and Technology–Hellas, Heraklion, 711 10, Greece. Email: fredyrapt@mailhost.ics.forth.gr.

George Papadopoulos, Institute of Computer Science, Foundation for Research and Technology–Hellas, Heraklion, 711 10, Greece. Email: grgepapadopoulos@gmail.com.

Nikitas Mavrakis, TIERRA Environmental, Heraklion, Greece. Email: nikitasmavrakis@gmail.com.

Michail Maniadakis, Institute of Computer Science, Foundation for Research and Technology–Hellas, Heraklion, 711 10, Greece. Email: mmaniada@ics.forth.gr.