Enhanced Plastic Recycling using RGB+Depth Fusion with MassFaster and MassMask R-CNN

Dillam Jossue Diaz Romero
Department of Mechanical Engineering & PSI-EAVISE- KU Leuven
Core Lab VCCM, Flanders Make
Leuven, Belgium
0000-0003-4309-3270

Isiah Zaplana
Department of Mechanical Engineering
Core Lab VCCM, Flanders Make
Leuven, Belgium
0000-0002-0862-3240

Simon Van den Eynde
Department of Mechanical Engineering
Core Lab VCCM, Flanders Make
Leuven, Belgium
0000-0001-9236-6162

Wouter Sterkens
Department of Mechanical Engineering & PSI-EAVISE- KU Leuven
Core Lab VCCM, Flanders Make
Leuven, Belgium
0000-0002-1675-0901

Toon Goedemé
PSI-EAVISE- KU Leuven
Sint-Katelijne-Waver, Belgium
0000-0002-7477-8961

Jef Peeters
Department of Mechanical Engineering
Core Lab VCCM, Flanders Make
Leuven, Belgium
0000-0003-1356-6508

***Abstract*— The rapid increase in waste generation from electrical and electronic equipment (WEEE) has created the need for more advanced sensor-based systems to sort this complex type of waste. Therefore, this study proposes a method for object detection, instance segmentation, and mass estimation of plastics and contaminants using the fusion of RGB and depth (D) images. The methodology is based on the Faster and Mask R-CNN with an extra head for the mass estimation. In addition, a pre-processing method to enhance the depth image (ED) is proposed. To evaluate the data fusion and pre-processing method, two data sets of plastics and impurities were created containing images with and without overlapping samples. The first data set contains 174 RGB images and depth (D) maps of 3146 samples, excluding their mass value, while the second data set contains 42 RGB and D images of 766 pieces together with their mass. The first and second data sets were used to evaluate the performance of Mask and Faster R-CNN. Further, the second data set was used to evaluate the network's performance with the additional head for mass estimation.**

The proposed method achieved 0.75 R^2, 1.39 RMSE, and 0.81 MAE with an IoU greater than 50% using the network Resnet50_FPN_RGBED. Hence, it can be concluded that the presented method can distinguish plastics from other materials with good accuracy. Furthermore, the mass of each detected particle can be estimated individually, which is of great relevance for the recycling sector. Knowing the mass distribution and the percentage of contaminants in a waste stream of mixed plastics can be valuable for adjusting the parameters of upstream and downstream sorting processes.

Keywords*— *artificial intelligence, plastic recycling, mass estimation, Mask R-CNN, Faster R-CNN, LiDAR, sensor-based material flow characterization

I. Introduction

While sensor-based sorting methods are widely adopted in state-of-the-art recycling facilities, sensor-based material flow characterization remains the subject of ongoing research [1]. Envisaged applications of sensor-based material flow characterization include automatic quality control, early detection of process failures, characterization of input flows, and data generation for adaptive process control [1] [2] [3]. The most promising results in sensor-based material flow characterization have been achieved in cases where Near-Infrared (NIR) has been combined with RGB and 3D cameras [4]. The use of cameras in addition to NIR is interesting, especially in cases where NIR typically has low performance, such as the distinction between bricks and concrete or the distinction between several black plastics [5] [6].

Some recent publications have presented practical implementations of sensor-based material flow characterization. In [7], the authors monitored the input flow, eject fraction, and reject fraction of a sensor-based (NIR) sorting process to evaluate the influence of the occupation density of the conveyor on the sorting performance. The significant relation between occupation density and sorting performance stresses the importance of adaptative process control to optimize sorting systems. Earlier research has demonstrated that sorting performances also depend on throughput, input flow composition, material shape, and volume flow fluctuations, which are characteristics that can all be quantitatively monitored in real-time using cameras [8] [9] [10]. The conveyor speed, the amplitude, and frequency of vibratory feeders, the screen cut, and shredder-related parameters, such as applied torque, shaft speed, and loading, are all examples of process parameters that could be adapted based on the collected data in real-time monitoring systems [10] [11] [12]. As such, sensor-based process control can contribute to establishing smart waste factories.

In addition to monitoring instantaneous characteristics of a material flow for adaptive process control, the extracted data can also be used to estimate the properties of the material flow as a whole. Recent research has demonstrated the possibility of forecasting the particle size distribution of a construction and demolition waste stream using a 3D camera and machine learning regression models [13]. Adapting the sorting process in function of the expected particle size distribution is especially relevant in sorting tasks where the particle size determines the quality of the sorted product, as is the case for stone material from construction and demolition waste that can be reused for, e.g., road construction.

However, the size distribution of particles in sorted fractions is often less important than the combined information of the material type and mass of the particles in the sorting outputs. Monitoring the material type and mass of all sorted particles would be the most direct way of evaluating the quality of most sorting outputs since purity targets of sorting outputs and tolerances for impurities are typically expressed in weight percent [14]. Furthermore, monitoring the particle masses and mass distributions of the input and output flows of density separation processes would be an excellent way of evaluating the performance of these processes, which are globally the most widely used method for sorting plastic waste streams due to their cost-effectiveness, high capacity, and relative simplicity of operation [15].

Therefore, the presented research developed a method to classify particles in a waste stream while simultaneously estimating their mass based on RGB+D images (color and depth). The method is demonstrated for a typical sorting task in the processing of Waste from Electrical and Electronic Equipment (WEEE), namely the distinction between plastics and contaminants [1] [16].

A. Related work for mass estimation in recycling

Estimating the mass of irregularly shaped objects can improve the assessment of the composition and purity of mixed waste streams containing both plastics and impurities. To the best of our knowledge, this is the first study to address simultaneous detection, segmentation, classification, and mass estimation of plastics and contaminants. However, previous research has used computer vision for mass estimation in distinct applications, such as recycling, medicine, agriculture, and robotics.

In 2018, an RGB image processing algorithm was proposed for calibration, background filtering, object detection, and mass estimation of plastic flakes [17]. The authors showed that the mass estimate helps to evaluate the flow of WEEE plastic flakes. However, the method did not allow the system to be used in real-world conditions since a fixed background and calibration are required for object detection. In 2022, the use of a convolutional neural network (CNN) for classification and a backpropagation neural network (BPNN) for mass estimation using up to 24 features extracted from depth images of metal scrap pieces was proposed [18]. However, this method does not use RGB images, which limits the detection and classification accuracy. In addition, an expensive line depth camera was required for classification, and object detection was based on a threshold that cannot be used if the material samples overlap.

The estimation method used in previous research can be used to determine the average density of an object class. However, it does not allow to overcome the issues with detecting overlapping samples or to improve classification accuracies. Therefore, the presented research uses data from RGB+D images to estimate the mass, object position, instance segmentation, and classification between plastic and contaminants by modifying the Mask Region-based CNN (Mask R-CNN) [19].

This study aims to develop a method for accurately estimating the mass of monitored plastics and contaminants, which in turn could be used for monitoring the density distribution of a material flow and thus provide feedback on the sorting efficiency of upstream processes. The study explores the benefits and limitations of using RGB and RGB+D images for mass estimation and proposes an additional output for the network, making the following contributions:

- Study of the feasibility of using a low-cost vision system consisting of a L515 RealSense LiDAR camera as a data source for mass estimation, object detection, and segmentation of plastics and contaminants. The research shows that the collected data can be used to estimate the mass of overlapping objects.
- The inclusion of an additional Region of Interest (ROI) head in the Mask R-CNN for mass estimation (MassMask R-CNN) opens the possibility of learning multiple mass distributions for each output class.
- Evaluation of RGB and D images fusion to obtain a more accurate mass estimation model for plastic and contaminants.

II. MATERIALS

Two data sets were developed in the presented study. Combined, these two data sets contain 3912 samples of mixed plastics and commonly encountered contaminants in WEEE. The collected WEEE samples have already been shredded and partially sorted. The material and class of each of the 3912 samples are determined and weighed individually.

The samples in each data set are divided into two main classes: plastics (65.1 wt%) and contaminants (34.9 wt%). These classes, in turn, contain several subclasses: the plastics class comprises black, white, and colored plastics, and the contaminants class comprises cables, printed circuit boards (PCBs), wood, metal, and unliberated plastics (plugs and sockets). The average mass of the plastic pieces in the data sets is 3.4g, with dimensions between 6 and 120 mm (about 4.72 in). The average mass of the contaminants is 4.9 g, with dimensions between 6 and 350 mm (about 1.15 ft).

Images were acquired and characterized using an illumination box for assuring a constant light environment, a scale for measuring the mass, and an Intel RealSense LiDAR L515 camera for RGB and depth imaging. The lightbox is a Caruba 70x70x70 with 6000 lumens dimmable to create a constant light environment. The system was configured to use the maximum intensity at 50% on the cool and warm dial. The pieces were weighed using an electric scale (Kern) with a resolution of 0.1g and a maximum load of 3000 g (about 6.61 lb). Regarding the camera, Intel introduced 2019 the L515 RealSense™ LiDAR, a time-of-flight depth camera [19] that captures RGB and depth (light detection and ranging) imagery. The Intel L515 has a depth resolution of 1024x768 with a depth angle of view (AOV) of 70° x 55° (horizontal x vertical (h x v)), a field of view (FOV) of 0.57m x 0.43m (h x v), and an RGB resolution of 1920 x 1080 with an AOV of 70° x 45° and a FOV of 0.57m x 0.32 (h x v).

Since the RGB and depth sensors have different FOV, the images must be mapped. The alignment process and image capture are performed simultaneously using the programming language Python 3.6 and the OpenCV, NumPy, and Pyrealsense 2 libraries. This process, predefined by Intel, reduces the original 1920 x 1080 and 1024 x 768 RGB images to 960 x 540 pixels and aligns them.

Once all aligned images were acquired, the ground truth (Class, Segmentation (Segm), Bounding Box (Bbox)) was generated using Hasty.ai, which provided a magic wand tool to segment an object within a region. Finally, the mass was added as an attribute per sample. Emailing the first or last authors can request access to the data set.

TABLE I. DESCRIPTION OF THE DATA SET WITH THE NUMBER OF IMAGES AND SAMPLES BY CLASS

	Data set 1			***Data set 2***	
Data set	***Class***	***Samples***	***Images***	***Samples***	***Images***
Train	Plastics	1378	120	348	28
	Contaminants	798		179	
Validation	Plastics	222	18	46	4

	Data set 1			Data set 2	
Data set	*Class*	*Samples*	*Images*	*Samples*	*Images*
Test	Contaminants	114	36	26	10
	Plastics	411		109	
	Contaminants	223		58	

As a result, data set 1 contains 174 RGB images and depth maps corresponding to 3146 samples, excluding the mass value in Table I. Data set 2 contains 42 RGB images and depth maps corresponding to 766 samples, including the mass value. For each data set, 50% of the images have no overlapping samples (Spread), while the remaining 50% were dropped from a fixed height resulting in some overlap and random positions (Overlapped), as shown in Fig. 1.

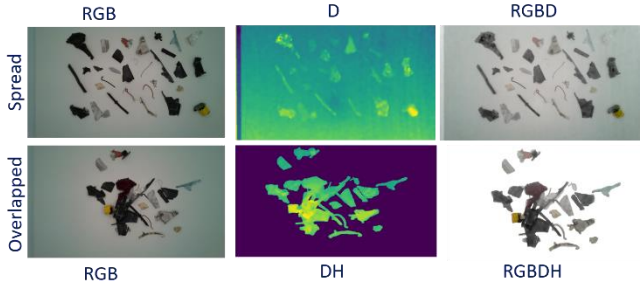

Fig. 1. The first row (Spread) shows an example of an image with plastics and contaminants with no overlap between the samples (left is an RGB image, the center is a depth image, and the right is an RGBD image). The second row shows an example of an image with overlapping plastics and contaminants (left is an RGB image, the center is an enhanced depth (ED) image, and the right is an RGBDH image).

As shown in Table I, the objects in each data set are randomly divided into 70% training, 10% validation, and 20% test objects for all the experiments. The experiments are computed using the GPUs: NVIDIA RTX3090 24 GB and NVIDIA TITAN 12 GB. ONE CPU: AMD Ryzen Thread Ripper 3990x 64-core with 128 GB DDR4 RDIMM memory. 16 bits depth image requires a pre-processing step to convert the images to 8 bits [18]. The first step is to define an area of interest (AOI) that extends from the surface where the object is placed to the highest depth intensity on the sample. Then, $v_{\max}$ and $v_{\min}$ are defined as the intensity values associated with the highest point of the samples and a point on the surface where the samples are located. In this case, the highest sample has a height of about 30 mm, corresponding to an intensity value of 1560, while a point on the supporting plate has a height of 0 mm, corresponding to an intensity value of 1350. This results in a z-axis resolution of 7 pixels/mm. Equations (1) and (2) are used to scale the image. $\mathrm{X} = \{x_{ij}\}_{i,j}$ denotes the point cloud matrix associated with the depth image. Such image is clipped according to equation (1):

$$\mathrm{X}_{\mathrm{scale}} = \left\{ x_{ij} \cdot \mathbf{1}_{[v_{\min} \le x_{ij} \le v_{\max}]} \right\}_{i,j} \quad (1)$$

where $\mathbf{1}_{\mathrm{cond}}$ is a characteristic function for condition cond, i.e., it returns 1 whenever the condition is satisfied and zero otherwise. Then, $\mathrm{X}_{\mathrm{scale}} = \{y_{ij}\}_{i,j}$ is normalized and transformed into an 8 bits image by virtue of equation (2):

$$\mathrm{X}_{8\mathrm{bit}} = \left\{ \frac{y_{ij} - v_{\min}}{v_{\max} - v_{\min}} \cdot 255 \right\}_{i,j} \quad (2)$$

An example of the resulting 8 bits image is depicted in the first row and second column in Fig.1 (DH). Although the camera is mounted with care to obtain a reliable image, the quality of the D-scan is not the best due to the uneven surface of the lightbox. As a result, the use of enhanced depth (ED) imaging is proposed, which uses the RGB image to remove the uneven background of the depth image (D), as shown in the second row and second column in Fig1.

The ED image is computed as follows: First, the RGB image is converted to a grayscale image and then converted once more to a binary image using the function cv2.threshold [20]. The following parameters were used to detect the image background, as well as the different objects: THRESH_BINARY_INV as the thresholding type, a threshold value (thresh) of 200, and a maximum value (maxval) of 181. The resulting binary image has a pixel value of 255 if the pixel is associated with one of the objects in the image and zero otherwise. Second, the information provided by the binary image is translated into $\mathrm{X}_{8\mathrm{bit}}$ (computed in equation (2)). For y_{ij} in $\mathrm{X}_{8\mathrm{bit}}$, the following correction is performed: $y_{ij} = y_{ij}$ if the pixel in the position i,j of the binary image has a value of 255, and $y_{ij} = 0$ otherwise. This results in a homogeneous depth image without background noise.

III. Methods

Mask R-CNN is an extension of Faster R-CNN that adds a branch for object mask prediction parallel to the existing branch for bounding box detection and classification. The versatility of the implementation of Mask R-CNN allows the generalization of other tasks, such as human pose and, in this case, mass estimation. As shown in Fig. 2, to evaluate the usefulness of the extracted data, Faster and Mask R-CNN have been modified in this study to input RGBD images while estimating the object's mass by adding a ROI head called MassMask R-CNN.

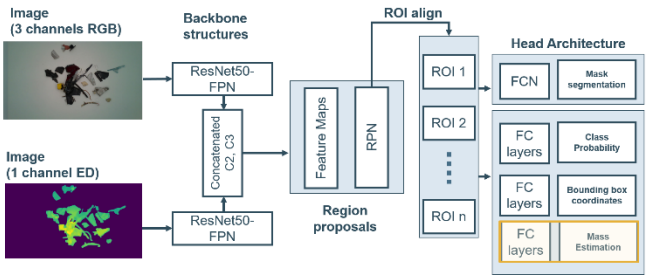

Fig. 2. Mask Region-based Convolutional Neural Network (R-CNN) architecture using a ResNet50-FPN backbone and additional mass estimation output (orange highlighted)[19], [21].

The MassMask R-CNN uses the exact structure of the Mask R-CNN, with one additional output. It has the object mask, a class label, an additional mass estimate, and a bounding box offset as outputs for each candidate object, like the Mask R-CNN.

Fig. 2 shows the flowchart of the proposed combined architecture. It consists of three primary building blocks: (1) the backbone structure of the network, adapted to input RGB or RGBD images, (2) the Region Proposals Network (RPN), without any modification; and (3) the Region of Interest Align (ROIAling). In addition, another building block is considered: (4) the head architecture, where an additional head is proposed for the mass estimation.

A. Backbone Structure

The first step is to input the RGB or RGBD images through a series of convolutional layers for feature extraction over the entire image. In this case, the proposed work utilizes ResNet with a depth of 50 [22] and a top-down architecture called Feature Pyramid Network (FPN), which was proposed in [22], and used in [19]. ResNet50_FPN consists of four convolutional modules (C1, C2, C3, and C4), and it is one of the most frequently used networks for feature extraction.

In the case of using RGB images as input, the backbone does not need to change. However, for RGBD fusion, the RGB and D images are fed into two separate subnetworks with the same architecture. Then, the parameters of layers C2 and C3 are concatenated in both subnetworks to obtain a uniform output [23]. In the following step, ResNet50_FPN_RGB, ResNet50_FPN_RGBD, or ResNet50_FPN_RGBED will denote either the ResNet50_FPN architecture for pure RGB images or the modified ResNet50_FPN for the fusion of RGB and D/ED images.

B. RPN

The result of the backbone structure is a set of three feature maps that serve as input to the Region Proposal Network (RPN). The RPN generates a set of bounding boxes or regions of interest (ROIs), each with an objectness score, i.e., the associated probability that there is an object in that box. The RPN is a fully convolutional network with an ordinary 3×3 convolutional layer architecture using nine anchor boxes and two 1×1 convolutional layers for regression and classification [21].

C. ROIAlign

Mask R-CNN has a particular layer, the ROIAlign layer, that has been designed to solve the lack of ROI alignment of the Faster R-CNN and its ROIPool layer. The ROIPool and ROIAlign layers homogenize the size of each ROI generated by the RPN so that it can serve as input to the following fully connected layers by using the sections of the feature map containing the homogenized ROI [19], [21]. While in ROIPool, quantization was used to correct the misalignment between the ROI boundaries and the feature map's boundaries, ROIAlign improves the ROI detection using bilinear interpolation.

D. The architecture's head

The network's head is known to be the last part of the network architecture, i.e., the one connected to the network's output. Faster and Mask R-CNN are usually conformed by fully connected layers and/or two to three convolutional layers for mask detection. In this work, the MassMask R-CNN contains two Faster R-CNN heads [21], a binary mask [19], and an additional head, a fully connected layer, to compute the mass estimation per ROI.

E. Experiments

To verify the relevance of RGB, RGBD, and RGBED for plastic and contaminant detection and segmentation, twelve experiments are presented in this paper: six experiments are performed with the first data set to evaluate Faster and Mask R-CNN. When appropriate, additional D or ED images are used as input. The other six experiments are performed with the second data set, which also evaluates the performance of Faster and Mask R-CNN with an additional head for mass estimation. Further, it is tested whether the additional head has any effect on the performance of the networks.

F. Implementation Details

Since a large data set was not available, a fine-tuning method was required for training. For both cases (RGB and RGBD image inputs), a pre-trained ResNet50_FPN in Pytorch on COCO Train2017 was used [24]. In the RGB and RGBD experiments, the trainable convolutional modules of the backbone were C3-C4 and all convolutional modules, respectively.

As proposed in [19], the ROI is considered positive if the intersection over the union (IoU) with a ground truth box is at least 0.5 and negative otherwise.

During the retraining, Random Sized BBox Safe Crop, Shift Scale Rotate, Random Brightness Contrast, Hue Saturation Value, Random Gamma, Gauss Noise, Gaussian Blur, Horizontal Flip, and Vertical Flip are applied as data augmentation methods using the Albumentations library [24]. Samples are not partially trimmed during data augmentation, and the structural sample object is always preserved.

The first six experiments' learning rate is set to 0.006, while in the case of MassMask R-CNN, the learning rate is set to 0.003. The Stochastic Gradient Descent (SGD) [25] with a momentum of 0.9 is used as the optimization method for both cases. In the first round of the experiments, 35 batches over 50 epochs were trained. In the second round, 10 batches over 30 epochs were trained.

The proposed architecture uses a multi-task loss on each sampled ROI, i.e., $L = L_{\text{mass}} + L_{\text{CLs}} + L_{\text{bbox}} + L_{\text{mask}}$. The sub-loss functions $L_{\text{CLs}}, L_{\text{bbox}}$ and L_{mask} are identical to those defined in [19], [21]. The L_{mass} is the Mean Absolute Error (MAE or L_1-Loss) and has been used in previous works for mass estimation [18].

G. Evaluation metrics

For the evaluation of R-CNN prediction in the proposed data sets, the COCO metrics were used [26], where the Average Precision (AP) and Average Recall (AR) are evaluated for the Bounding Box (bbox) and segmentation (segm) for a given IoU (50, 75, and 50:95). In the case of mass estimation, three different metrics, namely R-squared (R^2), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE) were used to evaluate regression performance.

IV. Results and Discussion

Table II compares the performance of Faster R-CNN and Mask R-CNN for the detection, classification, and segmentation of plastic and contaminant objects in data set 1. As shown in bold in Table II, the best object detection results are generally obtained by combining RGB and ED images, which show, in the case of Faster R-CNN, an increase of 10% in $AP_{50:95}$ when compared to RGB images. Previous work has shown that mid-to-late fusion of RGB and depth images can improve object recognition [23], [27]. For the Mask R-CNN, there is a slight difference between whether the RGB and RGBD or RGBED inputs are used. The RGB system results have marginally better AP_{75} and $AR_{50:95}$ for the bbox detection and segmentation.

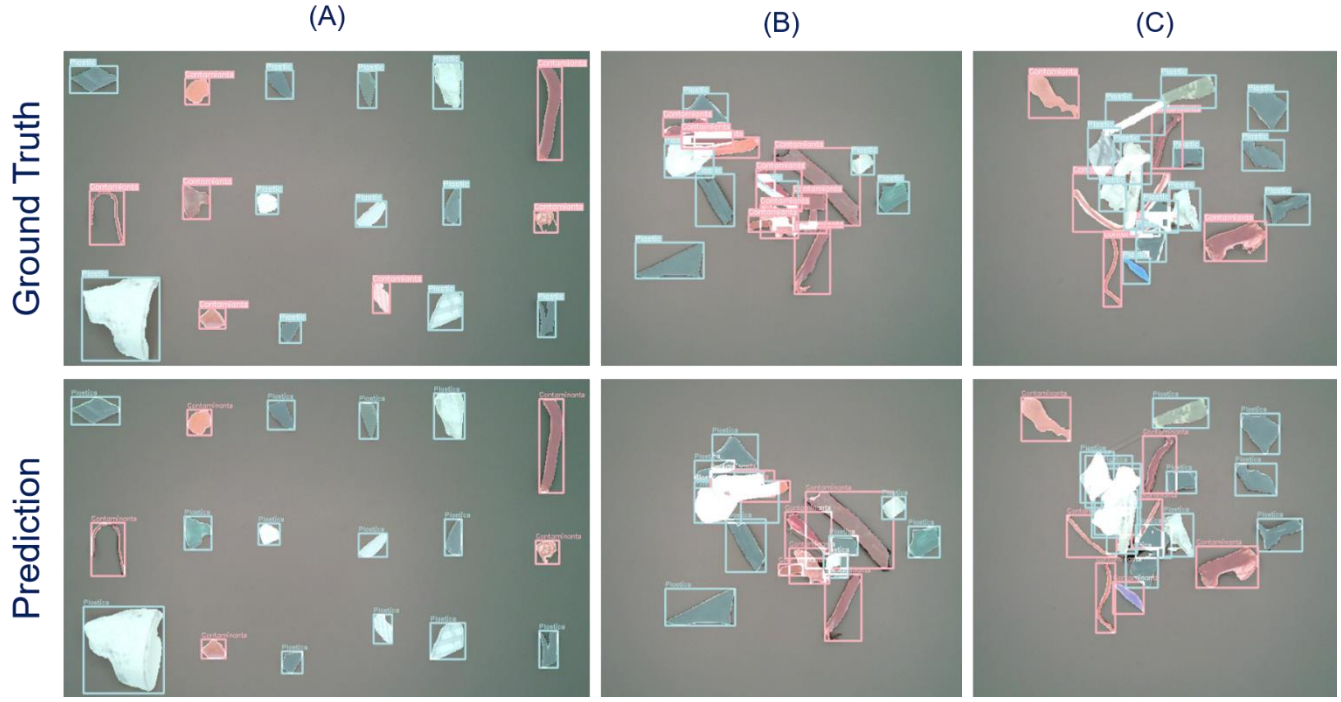

Fig. 3. Results of segmentation of plastics and contaminants with RGBED at a mask score greater than 55% and an IoU of 50%: (A) example of an image where the material samples do not overlap, (B) and (C) examples where the material samples overlap. The first row shows the ground truth, and the second the prediction, where the contaminant samples are colored in **light red,** while plastic samples are colored in **light blue**.

TABLE II. RESULTS DATA SET 1

		Faster R-CNN			
Type of Image	***type***	AP_{50}	AP_{75}	$AP_{50:95}$	$AR_{50:95}$
RGB	bbox	0.90	0.81	0.71	0.81
RGBD	bbox	0.91	0.81	0.77	0.77
RGBED	bbox	**0.90**	**0.82**	**0.82**	**0.81**
		Mask R-CNN			
Type of Image	***type***	AP_{50}	AP_{75}	$AP_{50:95}$	$AR_{50:95}$
RGB	bbox	0.90	0.82	0.72	**0.82**
	segm	0.83	0.69	0.61	0.74
RGBD	bbox	0.90	0.82	0.72	0.83
	segm	0.84	0.67	0.60	0.72
RGBED	bbox	**0.91**	**0.83**	0.71	0.81
	segm	**0.86**	0.68	**0.62**	0.74

TABLE III. RESULTS DATA SET 2 (MASS ESTIMATION)

		Faster R-CNN				Mass Estimation		
Channels	***type***	AP_{50}	AP_{75}	$AP_{50:95}$	$AR_{50:95}$	$R2_{50}$	$RMES_{50}$	MAE_{50}
RGB	bbox	0.78	0.71	0.61	0.80	0.57	1.62	1.0
RGBD	bbox	0.88	0.81	0.68	0.81	0.60	1.65	**0.9**
RGBED	bbox	**0.89**	**0.87**	**0.74**	**0.85**	**0.62**	**1.60**	1.0
		Mask R-CNN				Mass Estimation		
RGB	bbox	**0.89**	**0.85**	0.73	**0.83**	0.67	1.59	0.83
	segm	**0.90**	0.82	0.69	0.82			
RGBD	bbox	0.86	0.84	0.70	0.82	0.65	1.60	0.86
	segm	0.85	0.81	0.66	0.81			
RGBED	bbox	0.87	0.82	0.69	0.82	**0.75**	**1.39**	**0.81**
	segm	0.86	0.72	**0.70**	0.82			

As a result, the RGBD backbone may be redundant for mask detection, as it does not provide any significant improvement either with or without the use of the D images. Regardless of whether mass estimation is required, the use of a dual image input could limit the GPU's memory for processing multiple images simultaneously.

A traditional recycling company's capacity of sorting lines is about 5 tons/hour. The marginal difference of 1% AP between Resnet50_FPN_RGBED and Resnet50_FPN_RGB results in this context in an additional 50 kg of correctly classified material per hour when using the Resnet50_FPN_RGBED system. When the ratio of plastics to contaminants is 65% to 35%, this reduces around 17.5 kg of wrongly classified contaminants that end up in the plastic fraction. This amount is significant since plastic recycling is commonly only feasible if the plastic fraction has a purity above 97% [28].

Fig. 3. shows the mask R-CNN results for the best model, i.e., Resnet50_FPN_RGBED at a mask score greater than 55% and an IoU of 50%. This indicates that the selected model can easily classify most plastics (light blue) and contaminants (light red) with a given mask score. However, as shown in Fig.3. (A), there are two misclassifications. In this specific example, these misclassifications correspond to paper. This highlights one of the problems of the data set, which is the unbalanced number of contaminants. For example, there are only a limited number of paper contaminants in the original data set. Hence, it is difficult for the system to learn how to identify them correctly. It has also been found that, due to the camera's low resolution, the segmentation of small contaminants, such as cables, is difficult for the proposed

system. However, as stated in the introduction, this work focuses on vision-based classification and mass estimation of low-cost systems that can be easily implemented and integrated at different points of current sorting systems in recycling companies.

Fig. 3 (B) and (C) show the ground truth and the prediction for overlapping parts. The system has a considerably higher overprediction between the two classes in these two cases. A white mask is displayed over the object with multiple class detections, with an IoU greater than 50%. Increasing the IoU score to solve the problem in these cases is possible. However, the same IoU score is used for all visualizations to have a consistent representation.

In the case of overlapping objects in the image, Mask R-CNN performs better than Faster R-CNN in identifying objects, as shown in Table II. This stresses the added value of using Mask R-CNN and a vibratory feeder to distribute the parts, which improve the method's overall performance since Mask R-CNN can better handle objects that do not overlap.

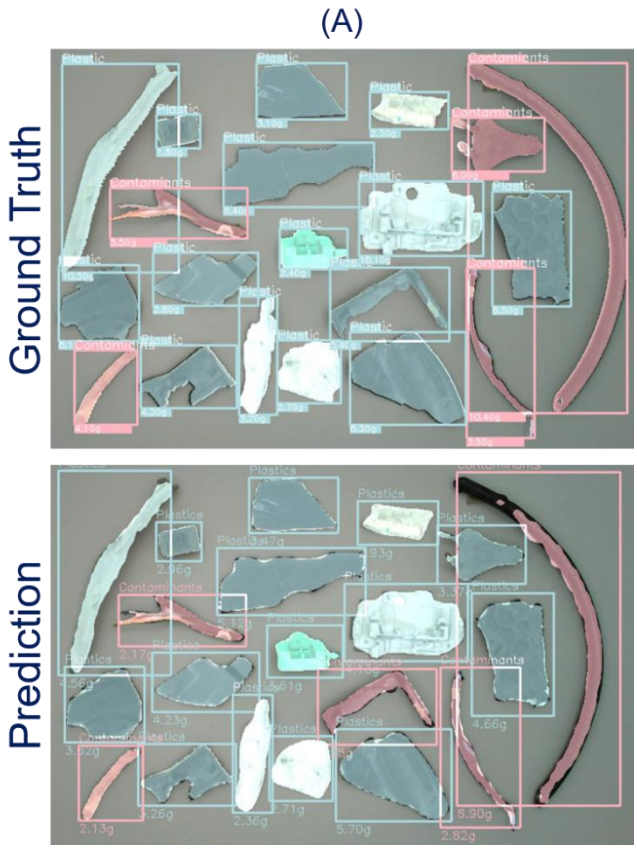

Fig. 4. RGBED mask for mass estimation at a mask score greater than 55% and an IoU of 50%. The top image shows the ground truth with the mass values at the bottom of the bbox of each object. The bottom image shows each object's class assigned and mass prediction values.

The results of the MassMask R-CNN on the second data set show that it is possible to predict the samples' mass regardless of using the D image, as shown in Table III. For MassFaster R-CNN, a fusion of the RGB and ED images was found to significantly increase $AP_{50:95}$ and $AR_{50:95}$ by 14% and 5%, respectively. This could be due to the fact that the learning of the mass is limited only by the input of the RGB image, which affects the overall performance of object recognition. In addition, using ED images instead of D images improves the location and mass prediction of the object by 6% for $AP_{50:95}$ and 4% for $AR_{50:95}$. The best mass estimation performance is obtained using Resnet50_FPN_RGBED with a dual backbone for feeding the data. The system has an R^2 of 0.62, RMSE of 1.6, and an MAE of 1 with an IoU above 50%, as shown in Table III.

As stated before, for MassMask R-CNN, the best segmentation performance is obtained using the RGB system only. However, the mass estimation is better when the RGBED model is used. This could be because the D images, while useful for mass prediction, are also redundant for segmentation.

Fig.4 shows the result of MassMask R-CNN for the RGBED case on an image from data set 2. In this image, two samples were misclassified (the socket in the top right corner and the L-shaped plastic in the middle right of the image). The L-shaped plastic is probably misclassified because this shape is more common among the contaminants. It is expected that this was misclassified for the socket because it contains both plastic and metal parts. The recycling company from where it was taken considers it a contaminant due to its metal content, which is why it has been labeled as such in the training set.

To solve these problems, it is necessary to create subclasses that allow the system to recognize different impurities or to define a specific class that predicts samples of plastic containing metal conductors (which are considered, by the recycling industry, as contaminants).

In addition, a confusion matrix was computed to evaluate the results of Resnet50_FPN_RGBED, as shown in Fig.5. The confusion matrix measures the performance in classifying plastics and contaminants. An additional row shows objects in the ground truth that were not detected, and an additional column shows objects that were not present in the ground truth. The confusion matrix shows that the system sometimes misclassifies contaminants, which could be due to the fact that parts are commonly considered contaminants even though they are made of plastic and other materials, leading to confusion in the predictions.

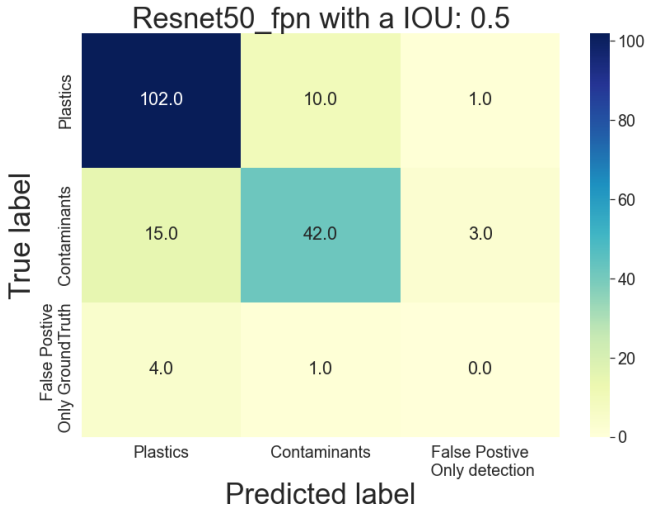

Fig. 5. Confusion matrix of Resnet50_FPN_RGBED using data set 2 for the mass estimation case.

For a better understanding of MassMask R-CNN for the Resnet50_FPN_RGBED case, Fig.6 shows the regression results used for mass estimation of all test patterns. Previous research has shown that the density and, thus, the object's mass can be more accurately estimated if the object class is

known, i.e., the network can learn the density per object type [18]. In this case, the mass estimation shows a trend by class with an R^2 of 0.75, RMSE of 1.39, and an MAE of 0.81 with an IoU of above 50%, as shown in Table III and Fig.6.

However, one of our system's limitations is the contaminants' generalization, which prevents the network from learning the density distribution per object subclass with the same material composition (e.g., cable, printed circuit boards (PCB), wood, metal, and plugs).

To improve the mass estimation, contaminants must be divided into subclasses, avoiding using a general class such as "contaminants". This was impossible in this study due to the limited number of samples in the data set for each one of the proposed subclasses. Nevertheless, the experiments show that using MassFaster and MassMask R-CNN is a new and promising alternative for mass estimation, segmentation, and classification of plastics and impurities. The proposed method could be adapted to unconventional materials and used as a monitoring system for the first, the intermediate, and the last step, hence, evaluate the performance during (pre-) sorting.

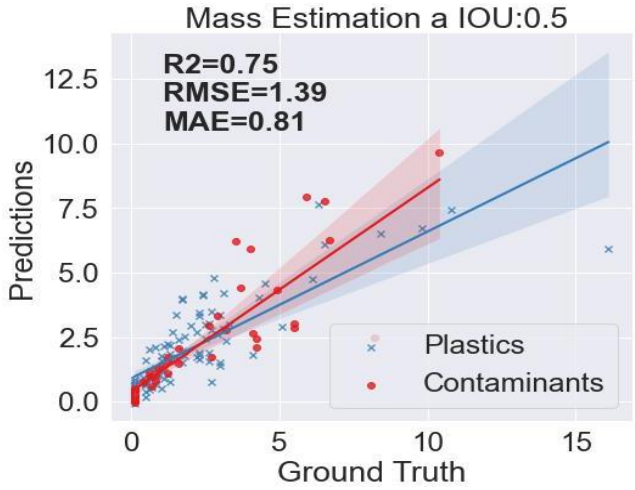

Fig. 6. Regression results using the Resnet50_FPN_RGBED architecture and including the R2, RMSE, and MAE metrics and the resulting regression lines with a 95% confidence interval for each regression (intended to show only the data trend).

V. INDUSTRIAL APPLICATION

The presented method to distinguish plastics from other materials and estimate the mass of each detected particle individually is of high relevance to the recycling sector. The following section describes an application where the proposed system could be used to serve a double function: on one hand, it could serve as a material flow characterization system to monitor the mass distribution of the particles in the material stream, and on the other hand, enhance an existing sorting step.

At a large Belgian recycling facility that was involved in this research, WEEE is processed as follows. After shredding, ferrous metals are removed from the waste stream with magnets, and non-ferrous metals are removed in a density separation process with a medium density of 1.4 kg/l. A NIR system sorts the floating fraction resulting from this density separation process to separate plastics from all other materials. Through a series of density separation processes, the plastics are further processed at different medium densities to divide the plastics into four density fractions A series of density separation processes further process the plastics at different medium densities to split the plastics into four density fractions, which is common practice in the plastics recycling industry [29]. The contaminant stream is also further processed in the recycling plant to valorize most of the materials in this stream. Fig.7 shows a schematic representation of these processes.

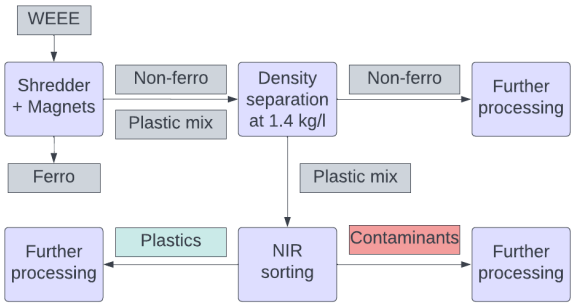

Fig. 7. Schematic of the sorting process, with the possible implementation of R-CNN's RGBED mask for mass estimation.

By implementing a camera system to monitor the material stream fed to the NIR sorting step, the performance of this sorting process could be enhanced. This article has proven that plastics can be separated from other materials based on camera images with relatively high accuracy with the presented classification approach. Since NIR technology has some inherent limitations, a combination of NIR and computer vision would achieve significantly higher sorting performances than the existing system. This should be the topic of future research.

Secondly, the mass estimation of particles in this waste stream would be of excellent value to enable feedback on the performance of the upstream density separation process. From experience, the recycling facility knows that the performance of their density separation processes fluctuates because it is hard to maintain the density of the separation medium at the desired level [30]. However, accurate real-time information on both the density of the separation medium and the performance of the processes is often still completely lacking. Knowing the mass distribution and proportion of contaminants in the plastics mix would be valuable in adjusting the sorting parameters of both upstream and downstream sorting processes. This is just one example of how sensor-based material flow characterization could improve sorting performances. However, it illustrates well the high relevance of this field of research and, in particular, the method presented in this article.

VI. CONCLUSIONS AND FUTURE WORK

The present research uses a Faster and Mask R-CNN for simultaneous mass estimation, object detection, classification, and segmentation of plastics and contaminants. To our knowledge, this paper is the first study combining object detection, mass estimation, and segmentation. This study investigates the benefits of using an additional architecture head in the Faster and Mask R-CNN for mass estimation. In addition, it identifies the system's performance in estimating the mass as well as in evaluating the fusion of RGB and depth images. The results obtained with Resnet50_FPN_RGBED show that objects can be classified with an AP and AR of up to 90% and an R^2 value of 0.75 for mass estimation. In future work, the data set will be enlarged and balanced to predict the contaminant subclasses and, thus, improve the system's accuracy, reducing misclassifications and significantly improving the mass estimation. In addition, the density variation of each impurity subclass will be investigated to

determine if it affects the performance of Mask R-CNN. This is also the reason why a low-cost vision system has been adopted in this study. A high-cost system with higher resolution could have allowed for better detection, classification, segmentation, and mass estimation results. However, it would have turned its implementation at several points of an industrial sorting system unfeasible. Finally, the system will be implemented at several points of a real-time sorting system to evaluate its robustness and to adjust the sorting parameters of the upstream and downstream sorting processes.

Acknowledgment

Thanks to F. Arslan for constructive discussions, revision, and support. Thanks to Abdalla Atef Abdelaziz and Basyiruddin Qori Wicaksono for generating the data set used in this research. This activity has received funding from the European Institute of Innovation and Technology (EIT), a body of the European Union, under the Horizon 2020, the EU Framework Programmes for Research and Innovation (project name: Increasing recycled content in added value products for resilient and digitized circular Economy (INCREACE), project number: 101058487)

References

[1] J. C. Hernández Parrodi *et al.*, 'Nahinfrarot-basierte Stoffstromüberwachung von Bau- und Abbruchabfällen', in *Mineralische nebenprodukte und abfälle 8*, Berlin, Germany, Sep. 2021, pp. 92–111.

[2] B. Küppers, S. Schloegl, G. Oreski, R. Pomberger, and D. Vollprecht, 'Influence of surface roughness and surface moisture of plastics on sensor-based sorting in the near infrared range', *Waste Manag. Res.*, vol. 37, no. 8, pp. 1–8, Jun. 2019, doi: 10.1177/0734242X19855433.

[3] S. Serranti, A. Gargiulo, and G. Bonifazi, 'Hyperspectral imaging based platforms for particulate solids characterization, inspection and quality control. Case studies: Application to polyolefins recycling', in *2011 3rd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, Lisbon, Portugal, Nov. 2011, pp. 1–4. doi: 10.1109/WHISPERS.2011.6080855.

[4] C. Vrancken, P. J. Longhurst, and S. T. Wagland, 'Critical review of real-time methods for solid waste characterisation: Informing material recovery and fuel production', *Waste Manag.*, vol. 61, pp. 40–57, Mar. 2017, doi: 10.1016/j.wasman.2017.01.019.

[5] S. P. Gundupalli, S. Hait, and A. Thakur, 'A review on automated sorting of source-separated municipal solid waste for recycling', *Waste Manag.*, vol. 60, pp. 56–74, 2017, doi: 10.1016/j.wasman.2016.09.015.

[6] H. Beel, 'Sortierung von schwarzen Kunststoffen mit Hyperspectral-Imaging-Technologie', in *Recycling und Rohstoffe, Band 10*, Berlin, Germany, 2017, pp. 175–191.

[7] N. Kroell *et al.*, 'Assessment of sensor-based sorting performance for lightweight packaging waste through sensor-based material flow monitoring: Concept and preliminary results', in *9th Sensor-Based Sorting & Control 2022*, Aachen, Germany, Apr. 2022, pp. 35–53.

[8] B. Küppers *et al.*, 'Influence of material alterations and machine impairment on throughput related sensor-based sorting performance', *Waste Manag. Res.*, vol. 39, no. 1, pp. 122–129, Jul. 2020, doi: 10.1177/0734242X20936745.

[9] B. Küppers, I. Seidler, G. Koinig, R. Pomberger, and D. Vollprecht, 'Influence of throughput rate and input composition on sensor-based sorting efficiency', *Detritus*, vol. 9, pp. 59–67, Mar. 2020, doi: 10.31025/2611-4135/2020.13906.

[10] A. Curtis, B. Küppers, S. Möllnitz, K. Khodier, R. Sarc, and L. Kandlbauer, 'Real time material flow monitoring in mechanical waste processing and the relevance of fluctuations', *Waste Manag.*, vol. 120, pp. 687–697, Nov. 2020, doi: 10.1016/j.wasman.2020.10.037.

[11] B. Küppers, S. Schloegl, N. Kroell, and V. Radkohl, 'Relevance and challenges of plant control in the pre-processing stage for enhanced sorting performance', in *9th Sensor-Based Sorting & Control 2022*, Aachen, Germany, Apr. 2022, pp. 17–34.

[12] A. Feil, E. Coskun, M. Bosling, S. Kaufeld, and T. Pretz, 'Improvement of the recycling of plastics in lightweight packaging treatment plants by a process control concept', *Waste Manag. Res.*, vol. 37, no. 2, pp. 120–126, 2019, doi: 10.1177/0734242X19826372.

[13] N. Kroell, P. Schönfelder, X. Chen, K. Johnen, A. Feil, and K. Greiff, 'Sensorbasierte Vorhersage von Korngrößenverteilungen durch Machine Learning Modelle auf Basis von 3D-Lasertriangulationsmessungen', presented at the 11. Wissenschaftskongress Abfall- und Ressourcenwirtschaf, Dresden, Germany, Mar. 2022.

[14] J. C. Hernández Parrodi, K. Raulf, D. Vollprecht, T. Pretz, and R. Pomberger, 'Case study on enhanced landfill mining at MSG landfill in Belgium: mechanical processing of fine fracions for material and energy recovery', *Detritus*, vol. 8, pp. 62–78, Dec. 2019, doi: 10.31025/2611-4135/2019.13878.

[15] M. Gent, M. Menendez, T. Javier, and I. Diego, 'Recycling of plastic waste by density separation: Prospects for optimization', *Waste Manag. Res.*, vol. 27, no. 2, pp. 175–187, Apr. 2009, doi: 10.1177/0734242X08096950.

[16] G. Bonifazi, R. Gasbarrone, and S. Serranti, 'Detecting contaminants in post-consumer plastic packaging waste by a NIR hyperspectral imaging-based cascade detection approach', *Detritus*, vol. 15, pp. 94–106, Jun. 2021, doi: 10.31025/2611-4135/2021.14086.

[17] H. Ramon, J. Peeters, C. Beerten, L. Antico, J. Duflou, and W. Dewulf, 'EVALUATING THE SIZE AND COLOUR DITRIBUTION OF MIXED PLASTIC RECYCLATES USING COMPUTER VISION', in *Conference Proceedings: Towards a Resource Efficient Economy-7th International Symposium and Environmental Exhibition*, 2018, pp. 1–7.

[18] D. J. Díaz-Romero *et al.*, 'Simultaneous mass estimation and class classification of scrap metals using deep learning', *Resour. Conserv. Recycl.*, vol. 181, p. 106272, 2022.

[19] K. He, G. Gkioxari, P. Dollár, and R. Girshick, 'Mask r-cnn', in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961–2969.

[20] A. Mordvintsev and K. Abid, 'Opencv-python tutorials documentation', *Obtenido Httpsmedia Readthedocs Orgpdfopencv-Python-Tutroalslatestopencv-Python-Tutroals Pdf*, 2014.

[21] S. Ren, K. He, R. Girshick, and J. Sun, 'Faster r-cnn: Towards real-time object detection with region proposal networks', *Adv. Neural Inf. Process. Syst.*, vol. 28, pp. 91–99, 2015.

[22] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, 'Feature pyramid networks for object detection', in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2117–2125.

[23] D. Díaz-Romero, W. Sterkens, S. Van den Eynde, T. Goedemé, W. Dewulf, and J. Peeters, 'Deep learning computer vision for the separation of Cast-and Wrought-Aluminum scrap', *Resour. Conserv. Recycl.*, vol. 172, p. 105685, 2021.

[24] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin, 'Albumentations: fast and flexible image augmentations', *Information*, vol. 11, no. 2, p. 125, 2020.

[25] I. Sutskever, J. Martens, G. Dahl, and G. Hinton, 'On the importance of initialization and momentum in deep learning', in *International conference on machine learning*, 2013, pp. 1139–1147.

[26] T.-Y. Lin *et al.*, 'Microsoft coco: Common objects in context', in *European conference on computer vision*, 2014, pp. 740–755.

[27] T. Ophoff, K. Van Beeck, and T. Goedemé, 'Exploring RGB+ Depth fusion for real-time object detection', *Sensors*, vol. 19, no. 4, p. 866, 2019.

[28] EJ. Bakker, PC. Rem, and N. Fraunholcz, 'Upgrading mixed polyolefin waste with magnetic density separation', *Waste Manag.*, vol. 27, no. 29, pp. 1712–1717, 2009, doi: 10.1016/j.wasman.2008.11.006.

[29] L. Strobl, T. Diefenhardt, M. Schlummer, T. Leege, and S. Wagner, 'Recycling Potential for Non-Valorized Plastic Fractions from Electrical and Electronic Waste', *Recycling*, vol. 6, no. 2, 2021, doi: 10.3390/recycling6020033.

[30] A. Eggers, J. R. Peeters, L. Waignein, B. Noppe, W. Dewulf, and M. Vanierschot, 'Development of a computational fluid dynamics model of an industrial scale dense medium drum separator', *Eng. Appl. Comput. Fluid Mech.*, vol. 13, no. 1, pp. 1001–1012, 2019, doi: 10.1080/19942060.2019.1663559.