XLI CILAMCE
XLI Ibero-Latin American Congress on Computational Methods in Engineering
16-19 November, 2020 FOZ DO IGUAÇU PARANÁ - BRAZIL

Recyclable Waste Classification Using a Deep Learning Vision System

Rafael Meneguelli[1], Daniel C. Cavalieri[1], Cassius Z. Resende[1]

[1]*Dept. of Control and Automation, Instituto Federal do Espírito Santo*
Rodovia ES-010, km 6,5. CEP: 29173 - 087, Serra, Brazil
rafaelmeneguelli01@gmail.com, daniel.cavalieri@ifes.edu.br, cassius@ifes.edu.br.

Abstract. One of the biggest environmental problems that humanity have been facing is the amount of waste generated. The disorderly growth of large cities combined with consumption and industrialization are substantially increasing the amount of solid waste generation. Recycling is an essential resource for sustainable development of societies whereas, as it reuses waste and decreases the accumulation of garbage. Brazil is one of the countries in the world where most generate solid urban waste and which has one of the largest numbers of people who separate this waste manually, many times in deplorable working conditions. Selective waste sorting basically consists of segregation of recyclable waste. However, when performed manually, the practice of segregation may not be followed evenly. The use of automated systems are an alternative to make the waste segregation more efficient and safer. This paper describes the use of a computer vision-based method to detect 4 main types of solid waste: glass, metal, paper and plastic. To classify and detect each type of solid waste was used a Convolutional Neural Network Based on the Mask Region (Mask R-CNN). The classification generates a mask and bounding box. After training the model using TrashNet dataset was achieved 0.80 for mAP. The use of those information can provide the position of the objects at the scene and robotic arms can make the automatic waste sorting.

Keywords: Mask R-CNN, Waste Sorting, Robotics, Recycling, Deep Learning.

1 Introduction

Rapid urbanization and industrialization have been generating at unprecedented rate solid waste materials. This fact is one of the biggest problems regarding the trash accumulation and sustainable development [1]. One of the best alternatives to deal with this problem is recycling. Recycling has the potential to reduce the amount of garbage gathered on municipal landfills, mitigating environmental impacts and lowering the production costs of materials. Nevertheless, the most of time, the waste sorting for the recycling process is made manually and may cause several injuries to the people that work with it. Thus, the quality and efficiency of the recycling process is still open to improvements [2]. In addition, new solutions in this area can bring social benefits to countries like Brazil, which is one of the largest producers of solid waste along with China and India (see Figure 1).

The selective garbage collection basically consists of the segregation and preliminary sorting of waste recyclable. However, when performed manually, the practice of segregation may not be followed evenly. With this, several automation systems were developed to solve the problem of garbage separation. These systems are divided between direct and indirect. The direct segregation of waste is based on the implementation of automatisms that aim to achieve the objective of separating the materials in view of their physical /chemical properties such as magnetic susceptibility, electrical conductivity and density. The indirect segregation, on the other hand, employs sensors to detect the presence and, the location of recyclable materials in the trash, so that machines or robots can be used to sort the recyclable materials detected[3].

The use of computational methods based on artificial intelligence as Deep Learning networks has the ability to make the separation process more efficient. In addition to the productive benefits, automatic garbage separation using resources like collaborative manipulator robots has the ability to provide several economic and social benefits[4]. As seen in [5] the problem of waste sorting could be solved using Near Infrared Images (NIR), this kind of image provides spectrometric characteristics of the objects in the scene, and the trash classification is made using classic statistical methods as Quadratic Discriminant Classifier (QDR) and geometric methods as Spectral Angle Mapper (SAM). In [6] deep learning methods was used to achieve the proposal. Knows networks as VGG16 and Resnet101 were used to classify the objects. After classification, the geometric center of the segmented objects was calculated and sent to a robot to separate different types of bottles.

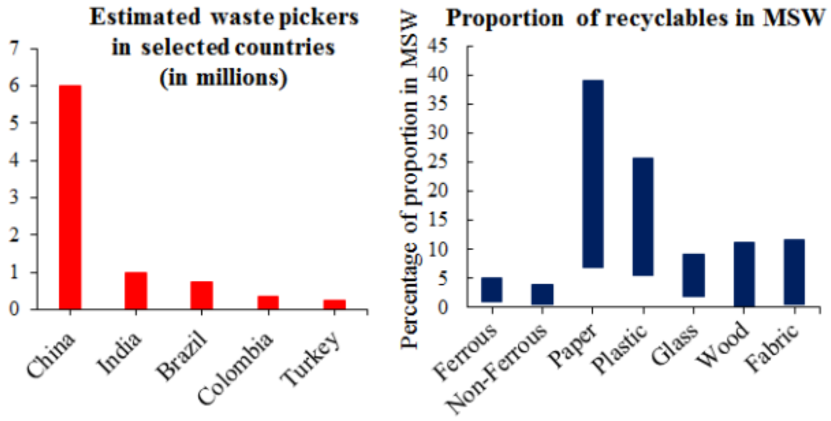

Figure 1. Amount of waste pickers in selected countries and principal type of MSW produced [1].

In this work a RGB camera will be used as vision sensor. The classification and training will be based on a deep learning network, the Mask R-CNN [7] that will be based on a ResNet 101 network pre-trained with the COCO Dataset database [8]. Furthermore, unlike the works cited, in the future, the separation of materials will be carried out by a collaborative robot using a simple interface. The vision system will communicate with the robot using Socket IP protocol, a computer will be responsible for process the data, classify the images, calculate the position of objects and send the position to collaborative manipulator.

2 System Overview

The present paper describes a system that can be considered as a prototype to an industrial garbage classification and segregation. Four types of common solid waste are going to be classified: glass, metal, plastic and paper.

A RGB camera is used to capture the raw images, those images are the Mask R-CNN model's input, once detected the objects, commands are sent to a robot to perform the waste separation. The Figure 2 shows the complete architecture of the system.

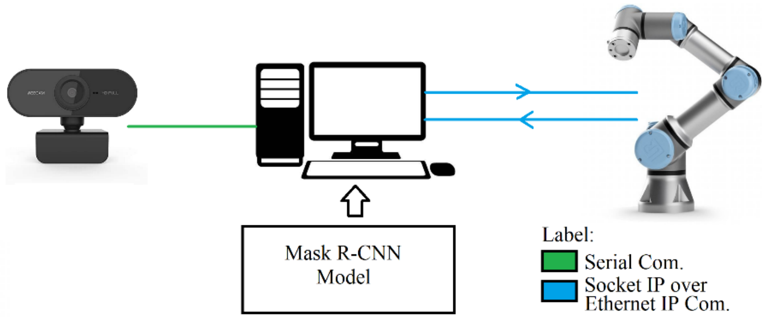

Figure 2. System architecture

This paper is going to describe only the computer vision system, responsible to detect the type of each recyclabe material. All the code was developed using python as programing language at the Google Colaboratory enviroment.

3 Data sets and proposed methods

3.1 Data set

As already mentioned in the last section, the proposed problem has four classes to segregate. Those classes were chosen because they are, according to [1] one of the most common type of garbage found in the domestic trash. Thereby, aiming to meet the needs of the problem the chosen dataset to train the network and solve the problem was the TrashNet data set [9]. This data set is composed by 6 classes of garbage: glass, metal, paper, plastic, cardboard and trash. Most of public datasets available presents those materials in original and undamaged state, however, actually those materials, disposed on the garbage, are fully dirt, damage and crumpled. The TrashNet data set portrays the real state of the recycle materials. The Figure 3 illustrates one example of each class of the data set used to train the proposed model.

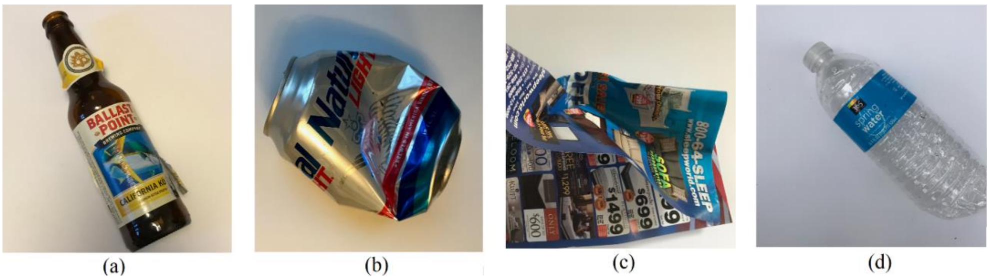

Figure 3. TrashNet dataset images: a) glass; b) metal; c) paper; d) plastic [9]

Each class of the data set has 400 to 500 images in different angles and levels of luminance which is good to make the model less sensible to light variations.

3.2 Preparing the Data set

To solve the problem of the classification of the recycling objects the data set was divided in train, validation and test. We separated to the train dataset 350 images of glass, 287 of metal, 415 of paper and 338 of plastic. The validation dataset is composed by: 126 glass images, 102 metal images, 148 paper images and 120 plastic images. Thus, test data set formed by 25 glass images, 20 metal images, 30 paper images and 24 plastic images. Thus, the data set was divided in 70% to train, 25% to validation and 5% to test.

To train and validate the Mask R-CNN model is necessary to annotate the target object regions on the data set images. It was used VGG Annotator Tool to define the regions of interest that were manually and precisely demarcated, as shown in Fig. 4. The result of the annotations generates a file in the .json (Java Script Object Notation) format which is responsible to store and transmit the information of labeling to the training model.

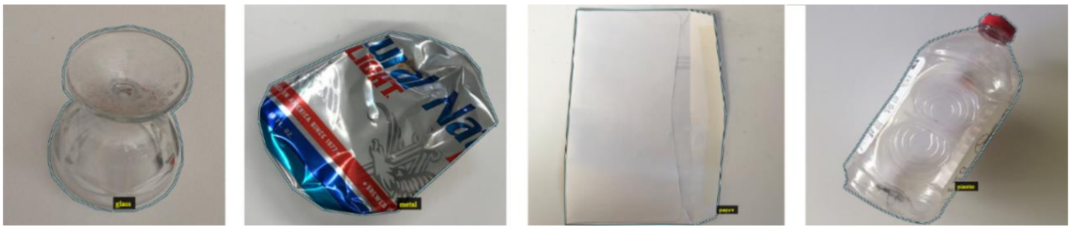

Figure 4. Annotated images of each class

Data augmentation technique is being used to increase the number of images used to train the model. The objective of this technique is to expand an existing training data base from manipulations in their shapes, contrast, brightness and translation. Some tests had been done using Gaussian noise and blur to the images in the data set augmentation process.

3.3 Mask R-CNN

Mask R-CNN (object detection model, whose term R-CNN comes from regional convolutional neural network) is a method that has the ability to segment object instances at the pixel level, in other words, this model is able to classify each pixel to the object to which it belongs. The result of the image classification is compound by a bounding box, class detection and the mask for each object in the scene [7].

There are two stages compounding Mask R-CNN model. First stage is basically responsible for generate proposals about the location of the objects in the input image scene. The second, is in charge to predict the object class, refines the bounding boxes and create the instance mask in pixel level of the objects detected on the first stage [10]. Those phases are both connected to the network backbone, in the case of this research the ResNet101.

The network backbone is basically a Feature Pyramid Network (FPN), structures used to detect and classify objects in different scales. Those Deep Neural Networks are responsible to make the feature extraction of the raw images and generate the pyramid map of the characteristics. Each level of the pyramid is connected to a branch of lateral connections responsible to realize the interaction between higher and lower levels of the pyramid. This kind of network makes a model scale-invariant, as it compensates the objects scale by the change of the pyramid level. This way, the backbone allows that a model detect objects in a wide variety of scales, making the classification robust and reliable. However, FPN backbones as ResNet 101, which generates a multi-level features pyramid has obvious limitation whereas the inference time increases according to the number of levels of the pyramid [11].

3.4 Waste classification using Mask R-CNN

Due to the high graphic processing to train a Mask R-CNN a transfer learning technique was used to improve the training results. This method consists of using a pre trained weights file from a different data set to get better performance of learning by avoiding much expensive data-labeling efforts. For this paper, were used the Microsoft COCO dataset [8] which, although not containing the classes used in this problem (glass, metal, paper and plastic), this data set contains approximately 120.000 other images with previous trained weights. This trained model has already learned several characteristics of several kind of objects that can be applied to detect and classify the recycle waste images in this research.

The Figure 5 illustrates the process flow used in this paper to perform the image classification of glass, metal, paper and plastic.

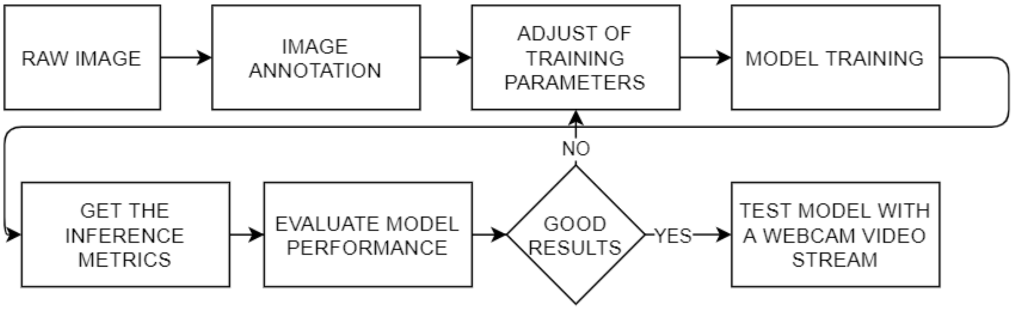
RAW IMAGE
IMAGE ANNOTATION
ADJUST OF TRAINING PARAMETERS
MODEL TRAINING
NO
GET THE INFERENCE METRICS
EVALUATE MODEL PERFORMANCE
GOOD RESULTS
YES
TEST MODEL WITH A WEBCAM VIDEO STREAM

Figure 5. Block diagram of the used algorithm

Different parameters were tested to achieve the expected results as learning rate, the resize of the input images, batch size and the number of epochs.

After performing each train the model's performance was evaluated through analysis of the confusion matrix, generated using the test images and the mean average precision (mAP) score. This metric is very common to evaluate object detection. To understand how this metric is calculated it's necessary to define some concepts. The first topic that is necessary to know to understand the mAP score is the Average Precision (AP) score. AP may be defined as the average value of the precision across all recall values. The second important concept in the mAP metric definition is the Intersection over Union (IoU), that consists basically in summarizes how well the ground truth object overlaps the object boundary predicted by the model. For COCO dataset we have 10 threshold values for the IoU metric, in each value the AP is evaluated. The mAP score can be defined as the mean of the AP score in each IoU threshold value [12].

To get the best model configuration those scores were evaluated and the train parameters were manually modified until we get the expected results.

The training and the validation process were done using the images from the TrashNet data set, as soon as good results have been achieved some tests have been done using a webcam to take photos of the objects in a mounted scene and inference the classes and masks using the trained model.

3.5 Tracking the objects

The next step to solve the proposed problem would be to classify and delimit the objects on a video stream, from the selection of the best model generated.

In this step one of the most important result to be evaluated was the time of inference, because as long as it takes to get the inference of the image as lower will be the detection frame rate of the system. Thus, in order to track objects in the video, each frame was inferred using the trained model. While the inference is made all the model's input information, coming from the video, is ignored. In other words, the longer the inference time, the smaller the amount of movement information is captured by the system.

4 Experimental Results and Discussion

4.1 Adjusting the model using the TrashNet

As described in the last section many model configuration parameters were tested, for each combination were evaluated the confusion matrix and the mAP score to determine the performance of the model in the inference of the test images on the TrashNet data set. The Table 1 list the training parameters defined for each model test.

The confusion matrixes obtained by the training of each model setup used to make the inference on the test data set are illustrated at the Figure 6.

Table 1. Tested model parameters setups

Model Setup	L. Rate	Augmentation	Epochs	Steps per Epoch	Input Size	Net. Layers
1	0,001	-	100	100	1024x1024	Heads
2	0,001	Rotation, scaling, crop, noise and blur	150	100	1024x1024	Heads
3	**0,001**	**Rotation, scaling and crop**	**200**	**70**	**512x512**	**Heads**
4	0,001	Rotation, scaling and crop	200	70	512x512	All
5	0,01	Rotation, scaling and crop	200	70	512x512	Heads
6	0,0005	Rotation, scaling and crop	200	70	512x512	Heads

The mAP values acquired for each setup model can be seen at the Table 2.

Table 2. mAP score values of each model setup training

Model Setup	1	2	**3**	4	5	6
mAP	0,59	0,62	**0,80**	0,28	0,00	0.49

The inference result of the third model setup in test data set samples are illustrated on the Figure 7.

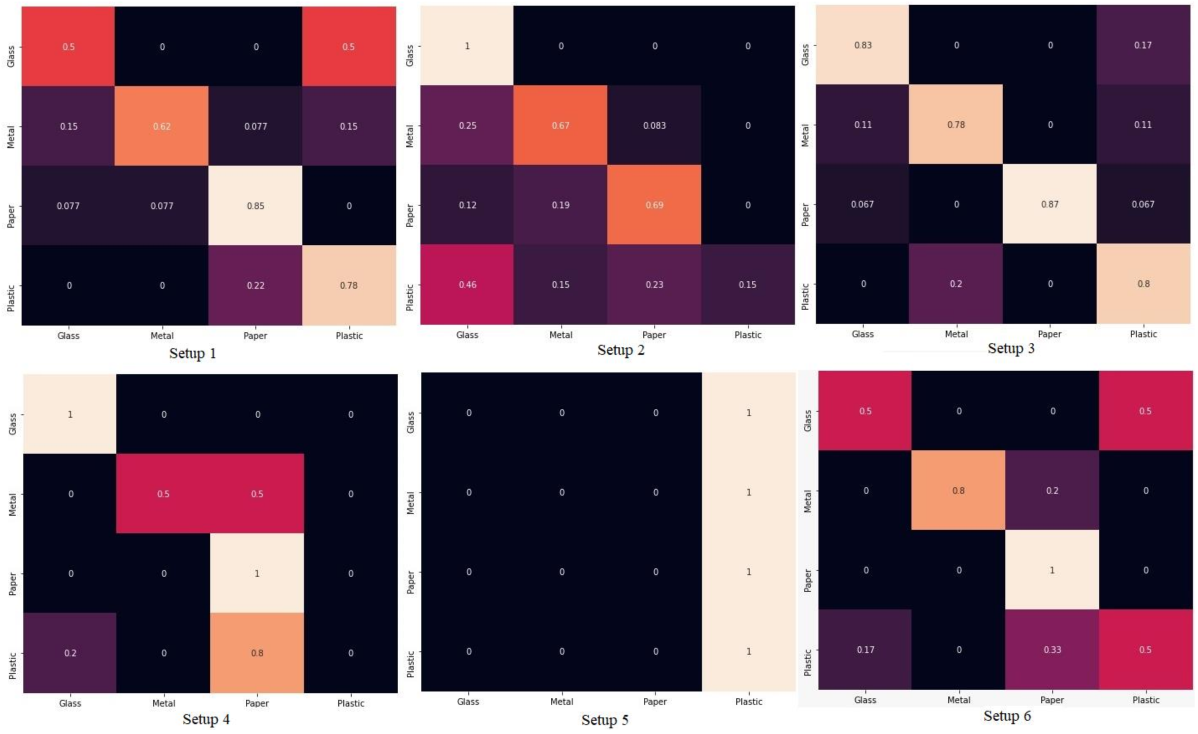

Figure 6. Confusion Matrixes obtained by the different setup models 1 to 6 respectively

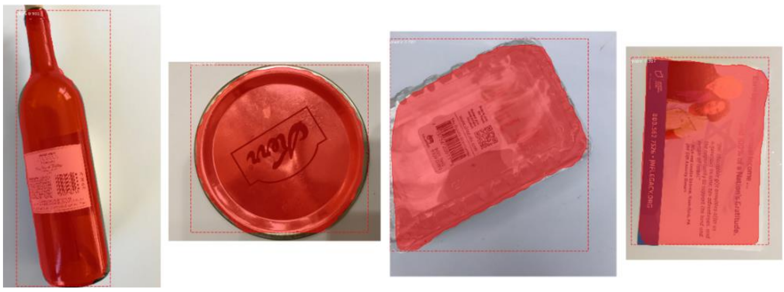

Figure 7. Instance segmentation presented by model 3 on test data set

Observing the confusion matrixes results another test was made, the glass class was removed from the data set and the model was trained using the same parameters as used at the model setup 3 with 3 classes, metal, paper and plastic. This class was removed because glass and plastic materials are very similar at the most of times because of their transparency. Performing the inference on the test data set the mAP score for the model with 3 classes was 0.95 and the confusion matrix is illustrated at the Figure 8.

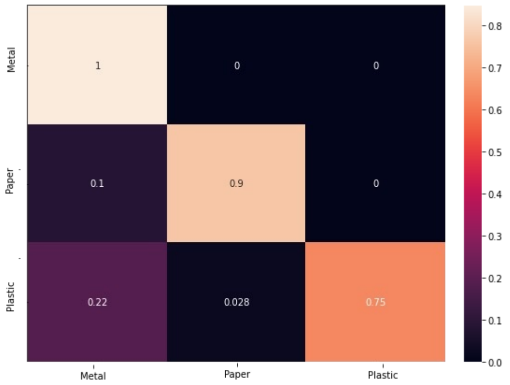

Figure 8. Confusion matrix for model trained without glass class using model setup 3

As observed model setup number 3 was the best with glass class included. Then, this model was tested making inferences in a video stream, captured by a webcam. The mean time of inference of each image using Google's Collab GPU was about 0.5 seconds, it means that the system can make the inference on 2 frames per second.

5 Conclusion

In this research the Mask R-CNN technique was used, with different training setup parameters to evaluate the performance of the model in solving the recycling waste classification problem. As can be seen from results, glass and plastic are visually similar materials and removing one of them of the data set improves considerably the performance and accuracy of the model in the garbage segregation.

As observed, the model setup number 3 proved to be the best parameter combination according to the confusion matrix and the mAP score obtained which was 0.80 with glass included and 0.95 without glass class. Testing this model with the chosen configuration with a webcam video stream it could be noted that the accuracy and the time of inference have been demonstrated enough to use this method to classify the recyclable waste and send the class, position and angle information to a collaborative robot to perform the waste separation.

For future work, we suggest to calculate the segmented object's position and angle in order to send by socket IP communication this information to a cobot. Another suggestion is to use a semantic segmentation like U-net trying to improve the inference time and system accuracy.

Aknowledgement

The authors thank FAPES – Fundação de Amparo à Pesquisa e Inovação do Espírito Santo, an agency of the State of Espírito Santo, Brazil, that supports scientific and technological development, for supporting this work.

Authorship Statement

The authors hereby confirm that they are the sole liable persons responsible for the authorship of this work, and that all material that has been herein included as part of the present paper is either the property (and authorship) of the authors, or has the permission of the owners to be included here.

References

[1] S. G. Paulraj, S. Hait, and A. Thakur, "Automated municipal solid waste sorting for recycling using a mobile manipulator," *Proc. ASME Des. Eng. Tech. Conf.*, vol. 5A-2016, no. September, 2016.

[2] A. C. Karaca, A. Ertürk, M. K. Güllü, M. Elmas, and S. Ertürk, "AUTOMATIC WASTE SORTING USING SHORTWAVE INFRARED HYPERSPECTRAL IMAGING SYSTEM Kocaeli University Laboratory of Image and Signal Processing (KULIS), MS MacroSystem Nederland ," *2013 5th Work. Hyperspectral Image Signal Process. Evol. Remote Sens.*, pp. 2–5, 2013.

[3] 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.

[4] A. H. Vo, L. Hoang Son, M. T. Vo, and T. Le, "A Novel Framework for Trash Classification Using Deep Transfer Learning," *IEEE Access*, vol. 7, pp. 178631–178639, 2019.

[5] A. Kulcke, C. Gurschler, G. Spöck, R. Leitner, and M. Kraft, "On-line classification of synthetic polymers using near infrared spectral imaging," *J. Near Infrared Spectrosc.*, vol. 11, no. 1, pp. 71–81, 2003.

[6] C. Zhihong, Z. Hebin, W. Yanbo, L. Binyan, and L. Yu, "A vision-based robotic grasping system using deep learning for garbage sorting," *Chinese Control Conf. CCC*, pp. 11223–11226, 2017.

[7] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2980–2988.

[8] T.-Y. Lin *et al.*, "Microsoft {COCO:} Common Objects in Context," *CoRR*, vol. abs/1405.0, 2014.

[9] M. Yang and G. Thung, "Classification of Trash for Recyclability Status," pp. 1–6, 2016.

[10] X. Zhang, "Simple Understanding of Mask RCNN," 2018.

[11] X. Li, T. Lai, S. Wang, Q. Chen, C. Yang, and R. Chen, "Weighted feature pyramid networks for object detection," *Proc. - 2019 IEEE Intl Conf Parallel Distrib. Process. with Appl. Big Data Cloud Comput. Sustain. Comput. Commun. Soc. Comput. Networking, ISPA/BDCloud/SustainCom/SocialCom 2019*, pp. 1500–1504, 2019.

[12] T. C. Arlen, "Understanding the mAP Evaluation Metric for Object Detection," 2018.