A New Paradigm for Waste Classification Based on YOLOv5

Mohammed SAJID[1], Nimali T MEDAGEDARA[2]

(1. *Department of Mechanical Engineering*, *University of West of England*; 2. *Department of Mechanical Engineering*, *Faculty of Engineering Technology*, *The Open University of Sri Lanka*, *Nugegoda*, *Sri Lanka*)

Abstract: Classification of garbage is of paramount importance prior to process them to categorise physically and this process helps to manage wastes by maintaining pollution free environment. Many systems that have capability segregate garbage are on the rise, but efficient and accurate segmentation with recognition mechanisms draw the attention of researchers. A computer vision approach for classifying garbage into respective recyclable categories could be one of the effective and economical ways of processing waste. This project mainly focused on capturing real-time images of a single piece of garbage and classifying it into three divisions: paper, or metal, or biodegradable (food) waste. Each garbage class contains around 2000 images obtained from an open-source dataset and images collected from Google and personally collected custom images. The developed intelligent models provide the effectiveness of the machine and deep learning in classification with structural and non-structural data. The model used was a Convolutional Neural Network (CNN) named YOLOv5. The project showcased vision based approach capable of maintaining an accuracy of 61%. The CNN was not trained to its maximum capacity due to the difficulty of finding optimal hyperparameters, as most of the images were gathered from Google Images.

Keywords: Waste Classification, YOLOv5

1 Introduction

Different strategies for garbage removal are in use. Dumping into others compound or isolated lands, thrown in the ocean, or burning with no concern of the environment are in fact common occurrences. Disposing of garbage is generally a significant problem for society.

Solid waste management is a global issue, which everyone is concerned in the world. It is the responsibility of any government to frame policies for waste management which is one of the main reasons attributable to well being of communities. Inefficient waste management cause pollution of oceans, blockages of drains and destroying flora and fauna. Moreover, the impact of improper waste disposal cause flooding during rainy season, transmitting diseases and increasing respiratory issues, such as airborne particles from incineration, harming animals that consume garbage, and affecting economic development that has an negative impact om tourism industry[1].

The globality of annual waste generation is on the verge of increasing and it is projected to reach 3.40 billion tones by 2050.Global waste will be developed by 70 percent on current levels by 2050. It has been estimated that around 40% of waste generated worldwide is not managed correctly instead of dumping and open incineration. A significant proportion of the total population has no access to appropriate waste disposal services, eventually it is poor waste mismanagement. There is even the problem of transportation of the trash due to rising transportation cost and this attributes to garbage collection around 39%[2] in low-income countries is. Improper garbage disposal devotes to disastrous epidemics of mosquito-borne malaria, hepatitis, and other potentially fatal disease[3], which endangers well being of society.

Every year around 2.01 billion metric tons of solid waste are generated [4]. This rubbish will put much pressure on the environment. If the garbage is not well classified, it will greatly harm the environment that the people live in. Classification would ease the waste management process.

1.1 YOLO Algorithm

For providing real-time object detection, YOLO algorithm is used with neural networks. The term YOLO means ‘You Only Look Once’. This algorithm has the ability to detect and recognize various objects as images in real-time. This algorithm is used in various applications with high speed and accuracy[5]. YOLO algorithm exploits convolutional neural networks (CNN) to identify images of objects in real-time. The models are trained to inspect the images and search for a subset of object classes. Object detection is a phenomenon in computer vision that identifies different objects in digital images or videos. The main advantage of real-time object detection models is that they are small and easy to operate by all developers.

1.2 Importance of YOLO algorithm

This algorithm has a predictive technique with minimal background errors and gives the exact outcome. This has good learning capabilities to learn the representations of objects and apply them in object detection [5]. This project utilizes YOLOv5s version of the YOLO model due to its faster speed, as shown in Fig.1.

Model	AP^{val}	AP^{test}	AP_{50}	$Speed_{GPU}$	FPS_{GPU}	params	FLOPS
YOLOv5s	36.6	36.6	55.8	**2.1ms**	**476**	7.5M	13.2B
YOLOv5m	43.4	43.4	62.4	3.0ms	333	21.8M	39.4B
YOLOv5l	46.6	46.7	65.4	3.9ms	256	47.8M	88.1B
YOLOv5x	**48.4**	**48.4**	**66.9**	6.1ms	164	89.0M	166.4B
YOLOv3-SPP	45.6	45.5	65.2	4.5ms	222	63.0M	118.0B

Fig.1 Different Version of YOLOv5 Model[6]

1.3 YOLOv5 Model

This model can be easily used, and the YOLOv5 framework is in custom domains. An object detector is accomplished to generate features from input images and then provide them through a prediction system to draw boxes around objects for predicting their domain[5].

One-stage detectors and two-stage detectors are two categories of object detection. Two-stage detectors decouple the task of object localization and classification for each bounding box. One-stage detectors make the predictions for object localization and classification simultaneously. YOLO is a one-stage detector (Fig.2); hence, You Only Look Once. Fig.3 depicts the basic anatomy of object detectors.

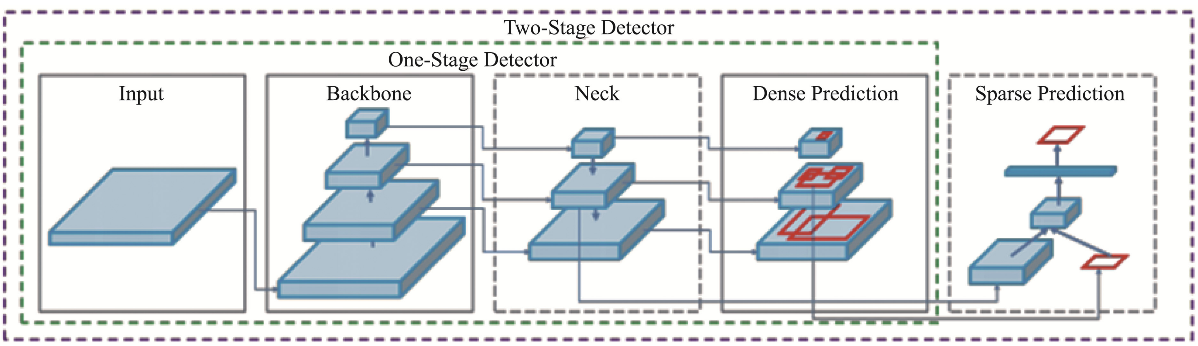


Fig.2 One-stage Versus Two-stage Detectors[5]

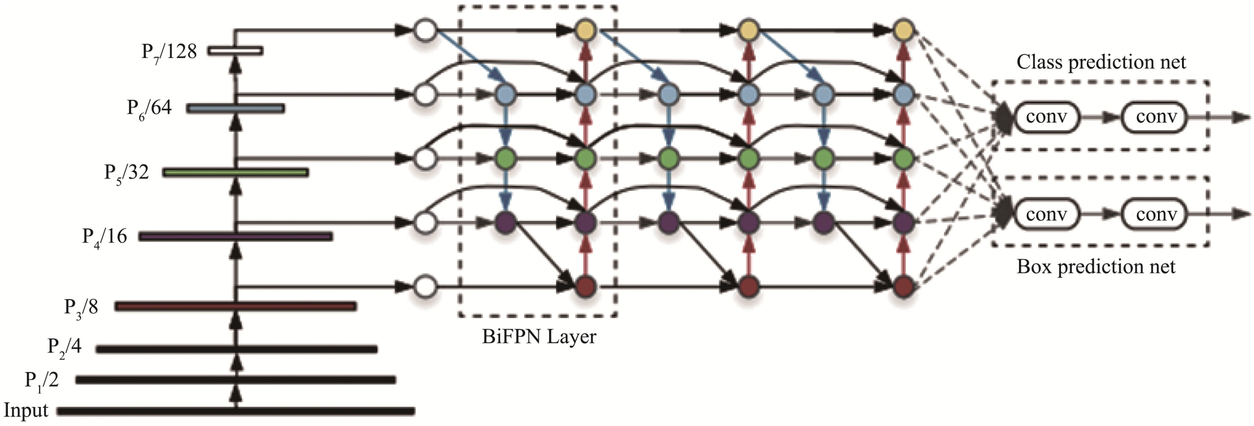


Fig.3 The Anatomy of an Object Detector [7]

1.4 YOLOv5 Model Architecture

Backbone, Neck, and Head are the three main stages of the YOLO [8-9]. Backbone is a CNN that combines and forms image features at different granules. Neck stage is a series of layers to mix and combine image features for prediction. Head stage absorbs features from the neck and captures box and class prediction.

1.5 Working of YOLO Algorithm

Working of YOLO algorithm follows three techniques such as residual blocks, bounding box regression and intersection over union[5-6].

1.5.1 Residual Blocks

Initially, the required image should be divided into number of grids. Grid division of the image as shown in Fig.4.

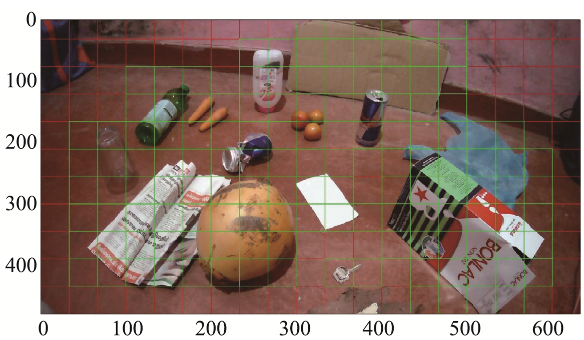


Fig.4 Grid Division of an Image

The Fig.4 shows the division of the input image into grids.

1.5.2 Bounding Box Regression

Outline that highlights an object in an image shown in Fig.5 is defined as a bounding box. Every bounding box in the image consists of the following attributes Width (bw), Height (bh), Class (c), Bounding box center (bx, by).

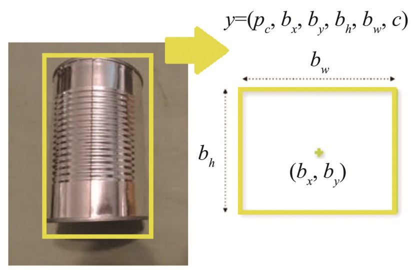


Fig.5 Example of Bounding Box Regression

1.5.3 Intersection Over Union (IOU)

IOU phenomenon is used in object detection. IOU uses in YOLO to permit an output box that surrounds the objects as shown in Fig.6.

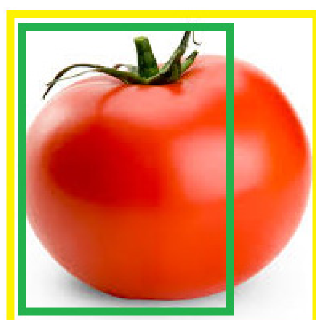


Fig.6 Example of IOU Working

2 Methodology

The project proposes a small-scale method to detect three different garbage classes: Biodegradable, Paper, and Metal. This could be implemented on a large scale for improved waste segregation.

3 Data Collection

Data collection is one of the major activities of training a model. A sufficient number of images are required in order for the model to work in high speed and with accuracy. The dataset for this project was obtained from various sources such as Open source data[10], images collected from Google sites and personally collected custom images were used for the research.

The dataset included in this study has 6000. Each class of trash has about 2000 images each. Out of which 70% of the images were used to train the model, 20% of the images were used to validate the model, and the rest 10% were used to test the model.

The next Step is labeling the images into the appropriate classes and checking if every image has its own .txt file.

3.1 Training & Validation

The data set was split into training and validation sets for validating the implemented model. The purpose of training set is to set the model in learning form. Validation set is used during training to iteratively estimate the model accuracy.

3.2 Testing

After training the model, the testing of the custom model was commenced. The working model can be used to detect different trash in pictures and videos. Testing was performed by individual image files being dragged onto the test area for the testing process. The model will return a confidence score for objects with the mean accuracy rate alongside that match the training data. Confidence scores will rise as the performance of the model improves through repeated training. Training a model can take several hours to complete. Retraining and testing of the model can be performed until the user is content with the model's performance. Fig.7 shows the testing and detection process while having a live feed from a Camera.

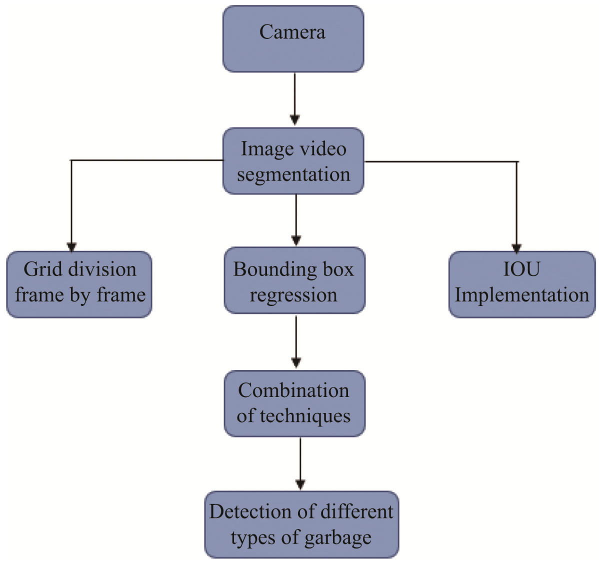


Fig.7 Flow Chart of Testing

4 Results

4.1 Individual Paper Testing

The Mean Average Accuracy (MAP) can be utilized for assessing YOLO object identification model. MAP compares the bounding box regression to the recognized box and returns the score. (AP) is the Average Precision and (IoU) defines the Intersection over Union, and this is used to measure the overlap between 2 boundaries. For measuring the accuracy of object detectors, AP metric is used.

Here are their mathematical definitions:

$$Precision=\frac{TP}{TP+FP} \qquad TP\text{=True positive}$$

$$TN\text{=True negative}$$

$$Recall=\frac{TP}{TP+FN} \qquad FP\text{=False positive}$$

$$FN\text{=False negative}$$

$$F1=2\cdot\frac{precision\cdot recall}{precision+recall}$$

In this datasets, IoU threshold was predefined as 0.5 for classifying whether the prediction is a true positive or a false positive.

Fig.8 shows testing done on Paper Waste. Table 1 shows the results achieved for paper waste.

Fig.8 Individual Testing: Paper Waste

Table 1 Results for Paper Waste

Precision	Recall	mAP at 0.5 IoU Threshold	mAP at 0.95 IoU Threshold
0.03	0.06	0.02	0.01

4.2 Metal Waste

Fig.9 shows the individual testing performances on metal waste. The following are the results achieved (Table 2).

Fig.9 Individual Testing: Metal Waste

Table 2 Results for Metal Waste

Precision	Recall	mAP at 0.5 IoU Threshold	mAP at 0.95 IoU Threshold
0.77	0.54	0.65	0.39

4.3 Biodegradable

Fig.10 shows testing done on Biodegradable Waste. The following are the results achieved (Table 3).

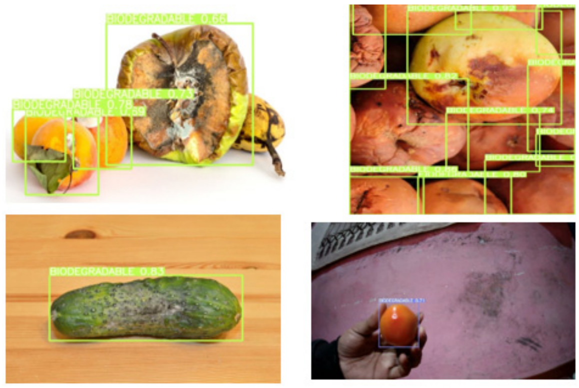


Fig.10 Individual Testing: Biodegradable Waste

Table 3 Individual Results for Biodegradable Waste

Precision	Recall	mAP at 0.5 IoU Threshold	mAP at 0.95 IoU Threshold
0.83	0.45	0.58	0.30

4.4 Batch Testing

The Table 4 shows the results achieved for overall model performance by testing of model concurrently in a batch.

Table 4 Results for Batch Testing

Precision	Recall	mAP at 0.5 IoU Threshold	mAP at 0.95 IoU Threshold
0.91	0.61	0.50	0.30

4.5 Mosaic Augmented Batch Testing

Fig.11 shows implementation of the Data Augmentation method known as Mosaic Augmentation. This merges four training images into one in certain ratios. This allows for the model to learn how to identify the objects when cropped, or different rotation, or flipped, or blurred or at a lower scale than normal[8-12]. It also is useful in training to significantly reduce the need for a large mini-batch size[8]. The following are the numbering system followed in order to identify the object according to the available classes: 0 - Biodegradable, 3 - Metal, 4 – Paper.

Fig.11 Mosaic Augmented Testing

Raspberry Pi running the latest version of Raspbian OS was used for the purpose of real time identification. Python 3 was used as the development environment. After capturing the objective image and the reference image is identified, the rest of the process was completely automatic. Hence no need for user intervention. The algorithm has been applied to the complete image.

4.6 Overall Results

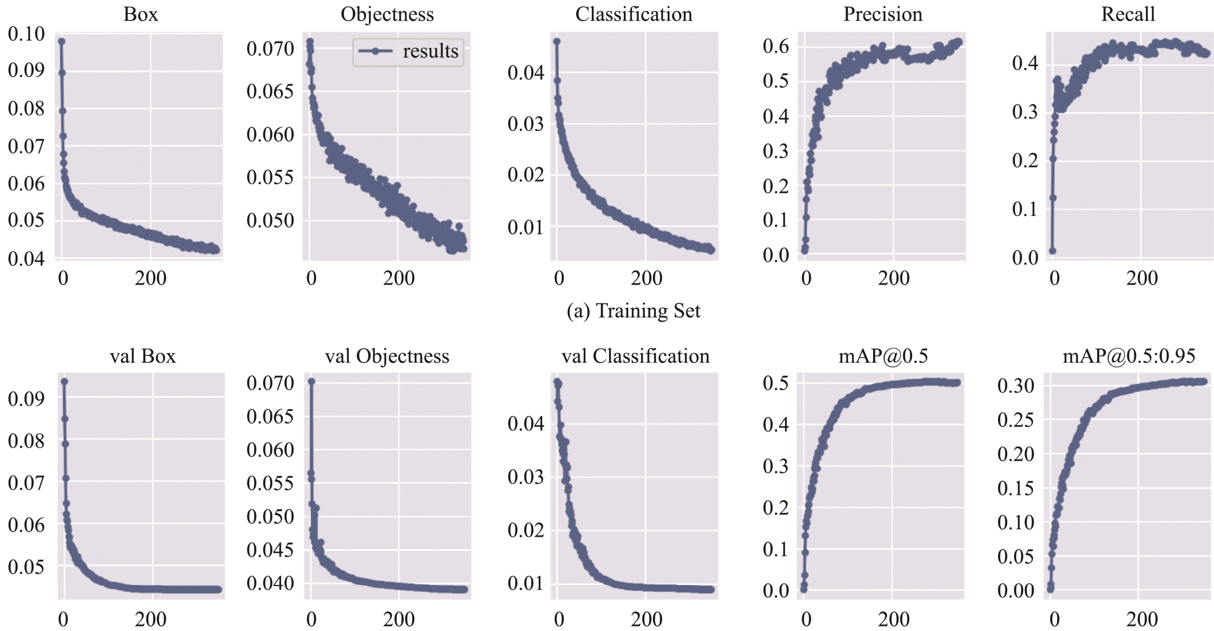
In the results shown in Fig.12, x-axis indicates number of epochs and y-axis indicates model performance. The plots Fig.12 (a) include bounding box loss, objectness loss, classification loss over the training epochs for the training set. Fig.12 (b) includes the validation set of bounding box loss, objectness loss, classification loss, precision, recall and mean average precision (mAP) at 0.5 and 0.95 IoU threshold over the training epochs.

4.7 Model Performance Analysis

The epoch accuracy is used to identify the model's trained quality and its performance for future prediction. The graph (Fig.13) represents Performance Value (y-axis) versus Number of Epochs (x-axis). The model was run for 350 epochs and trained for 13 hours. The model performance graph (Fig.13) shows accuracy and loss metrics of Training and validation sets. It can be clearly seen that accuracy of both sets increases over time and the reach saturation points at around 250 epochs. And the loss for both sets remains low which is essentially good for the model.

This research was conducted to identify how well the model was able to detect real – time trash and classify them into different categories. The experiment was performed by placing different trash (paper, metal, biodegradable) in front of the camera which was previously not added to the dataset. The results showed 61 % mean accuracy towards most images. The analysis was done by using 6000 images and due to this large dataset, each of the features in the images could be properly analyzed during the modelling process.

When the TensorFlow Lite (TFlite) model was running 1.07 FPS (Frame Per Second) was achieved. Therefore, these results show that a mobile detector and/or a Smart Garbage Bin with YOLOv5 can be used for real-life simple trash identification and collection.



(a) Training Set

(b) Validation Set

Fig.12 Overall Results

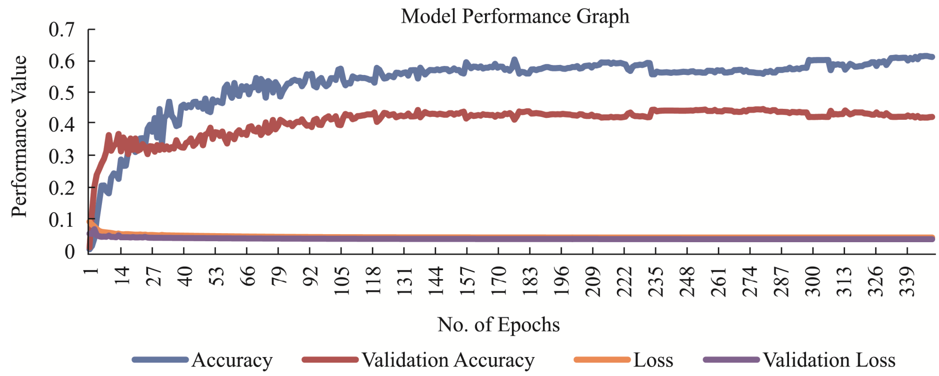


Fig.13 CNN Model Performance Analysis

4.8 Problem with Paper

Among the 2500 images present in the Paper Class, most of the images taken for training the model, consisted of zoomed in image of paper (Fig.15). This was developed due to a random function implemented in order to keep the train set, validation set and the test set neutral. But due to this function the accuracy of detecting paper was decreased. Fig.14 shows the Overall Heatmap of all the images in the paper class, it can be seen that the annotation for paper went from end to end, that is, from y_{max} to x_{max}. But the model still accurately detects paper under proper lighting.

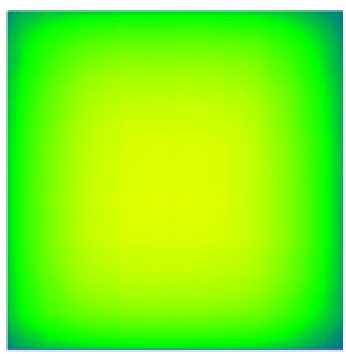


Fig.14 Annotation Heatmap of the Paper Class

Fig.15 Error in Images of Paper in Dataset

5 Conclusion

The aim of this project was to investigate the suitability of running a real - time object detection system on a Raspberry Pi and on Google Colab Pro to identify different types of trash. YOLOv5s and Tflite models were implemented and tested for accuracy and speed at different input sizes. The results showed that the model is accurate and that only in applications that do not require high speed would it be viable to use the Raspberry Pi as hardware. However, detections made on Google Colab were found to be faster and more accurate due to its running on a GPU. There is a trade-off with accuracy to be made if higher speeds are to be achieved since there is not enough computational power to have both. This study could be of help for others who will be looking to implement object detection/classification models on Raspberry Pi to make Garbage Detection projects of their own. The classification of trash into various recycling categories is possible through machine learning and computer vision algorithms. One of the biggest pain-point is the wide varieties of possible data. Therefore, in order to create a more accurate system, there needs to be a large and continuously growing dataset. The developed intelligent models provide the effectiveness of machine learning and deep learning in classification with structural and nonstructural data. The model used was a Convolutional Neural Network (CNN) named YOLOv5. The project showcased that the garbage classified was mostly accurate (61%); however, the CNN was not trained to its full capability due to the difficulty in finding optimal hyperparameters as most of the images were collected from Google Images. As further improvements author suggests to perform more testing of different trash, which comes in different sizes, shapes and color and to test the impact of lighting that has on a model's ability to detect trash. The future work can be focused to explore the on the above.

References

[1] Anon, *Trends in Solid Waste Management*. Available at:https://datatopics.worldbank.org/what-a-waste/trends_in_solid_waste_management.html [Accessed November 24, 2021].

[2] Ellis, C., 2018. World Bank: Global Waste Generation could increase 70% by 2050. *Waste Dive*. Available at: https://www.wastedive.com/news/world-bank-global-waste-generation-2050/533031/ [Accessed November 24, 2021].

[3] Kaza, S. et al., 2018. *What a waste 2.0: A global snapshot of Solid Waste Management to 2050*, Washington, DC: World Bank Group.

[4] Anon, 2021. Global waste index. *Sensoneo*. Available at: https://sensoneo.com/global-waste-index/ [Accessed November 24, 2021].

[5] Bochkovskiy, A., Wang, C.-Y. & Liao, H.-Y.M., 2020. *YOLOv4: Optimal Speed and Accuracy of Object Detection*. dissertation.

[6] Anon, Introduction to Yolov5. *YOLOv5 Documentation*. Available at: https://docs.ultralytics.com/ [Accessed November 24, 2021].

[7] Tan, M., Pang, R. & Le, Q.V., 2019. *EfficientDet: Scalable and Efficient Object Detection.*

[8] Solawetz, J., 2021. Breaking down Yolov4. *Roboflow Blog*. Available at: https://blog.roboflow.com/a-thorough-breakdown-of-yolov4/ [Accessed November 24, 2021].

[9] Solawetz, J., 2021. Yolov5 new version - improvements and evaluation. *Roboflow Blog*. Available at: https://blog.roboflow.com/yolov5-improvements-and-evaluation [Accessed November 24, 2021].

[10] Thung, G. & Yang, M., 2016. *Classification of Trash for Recyclability Status*.

[11] Aysegul TakimogluAysegul has experience in HR management & *, N., 2021. What is Data Augmentation? techniques, benefit & examples. *AIMultiple*. Available at: https://research.aimultiple.com/data-augmentation [Accessed November 24, 2021].

[12] Solawetz, J., 2020. Data augmentation in yolov4. *Roboflow Blog*. Available at: https://blog.roboflow.com/yolov4-data-augmentation/ [Accessed November 24, 2021].

Author Biographies

Mohammed SAJID is currently a B.Eng. candidate in Mechanical (Mechatronics) from University of West England. His main research interests include Computer Vision (CV) and Artificial Intelligence (AI).

E-mail: affilian147@gmail.com

Nimali T MEDAGEDARA received B.Sc. (Hons) degree in Mechanical Engineering from the University of Peradeniya (Sri Lanka) in 1993, and MPhil in Materials and Manufacturing Engineering from Sheffield Hallam Uni-University (UK). She is a charted engineer and member of the Institute of Engineers Sri Lanka (IESL), Member of IEEE and Member of SLAAS. She is now a senior lecturer in the open university of Sri Lanka, and the principal investigator in the Soft Robotics Research Group (SORRG) (https://www.ou.ac.lk/SRRG/) of OUSL. Her main research interests include soft robotics, finite element analysis and multi axial fatigue.

E-mail: tmmed@ou.ac.lk