IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⋇ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.11503

A Real-Time Application for Waste Detection and Classification

Minh Nguyen[1], Huy Lam[2], Tuan Le[3], Nha Tran[4], Tai Lam[5], Tinh Nguyen[6], Hung Nguyen[7]

University of Education, Ho Chi Minh city, Viet Nam[1, 2, 3, 4, 5, 6, 7]

Abstract: Nowadays, we are facing many problems of environmental pollution. One of them is the process of waste management since the amount of waste is proportional to the number of people in urban areas. The classification of waste plays an important role in the recycling of waste contributing to minimizing the risk of spreading pathogens, toxic and dangerous elements. We are in the fourth industrial revolution, applying cutting-edge technology is the trend, specifically deep learning techniques in the waste recycling process. Smart waste recognition also contributes to saving human resources and reducing costs for waste collection and recycling. In this paper, we propose a waste detection and classification model based on YOLOv4 architecture. We experimented and obtained mAP of 90.27%, F1-score of 86% on the dataset that we synthesized including 4 main types of waste: plastic, metal, glass, and paper.

Keywords: Computer Vision, Object Detection, Classification, Deep Learning, YOLOv4.

I. INTRODUCTION

Human society is developing constantly and has a modern economy. Besides, the consumption in society is very large and through that, the amount of waste produced has also increased to an alarming level. However, there are still kinds of waste that can be recycled. Classifying waste takes a lot of time and manpower. Therefore, in order to save time and costs, we propose a method to support waste detection and classification with the desire to help limit the negative impacts on the environment that waste causes. In recent years, many researchers have used deep learning models to detect objects, such as waste, and have had a lot of success. Awe et al [1] proposed an experimental method for classifying waste into three categories: recyclable, paper, and landfill based on the Faster R-CNN model, the authors achieved an accuracy of 68%. Two researchers Shanshan Meng and Wei Ta Chu [2] used models like DenseNet121, MobileNetV2, ResNet50… These authors experimented on the same dataset to show that the ResNet50 model has higher accuracy of waste classification; it would be improved from 91.40% to 95.35% with data augmentation. Berardina De Carolis et al [3] used the YOLOv3 model for real-time waste detection, the networks were fine-tuned to fit the collected dataset. Mittal et al [4] in SpotGarbage – a garbage detection application using deep learning, the authors used a pre-trained AlexNet model and achieved an average accuracy of 87.69%.

In this paper, we introduce a method to detect and classify four kinds of waste in real-time including plastic, metal, glass, and paper. We use deep learning techniques, specifically the YOLOv4 model [5] as the main framework to solve this problem. YOLOv4 is a Convolutional Neural Network model for object detection, recognition, and classification, it is created from the combination of convolutional layers and fully-connected layers. Convolutional layers extract the features of the image. Fully-connected layers predict the probability and the coordinates of the object.

The YOLO model is an algorithm used for real-time object detection that can be used for real-time image detection. Based on the research done by Juan Du, object detection based on CNN family and YOLO [6], it shows that the performance of the YOLO model is better than other CNN algorithms, such as Faster R-CNN. In the study of comparing versions of the YOLO model, Shuo Wang [7] compared the models YOLOv3, YOLOv3-tiny, YOLO v3-SPP3, YOLOv4, and YOLOv4-tiny. The results showed that when tested on the set Pascal VOC data, the mAP of the YOLOv4 model reached 87.48% (the highest), 14.2% higher than the YOLOv3 model. YOLOv4 meets the requirements of real-time and high accuracy.

II. METHODS

Classifying waste in an image is a problem of great practical significance to help reduce pollution and reduce the amount of waste. To be able to detect and classify waste automatically, we use the YOLOv4 model because of its high accuracy and ability in real-time conditions. The process of applying the YOLOv4 model to the problem of waste detection and classification will be divided into three stages:

- **Input:** Using waste images that have been annotated according to four classes of plastic, metal, glass, and paper.
- **Processing:** Training the YOLOv4 model to detect the position and classify each waste in the image.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⋦ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.11503

- **Output:** After the image goes through the model, it will produce the classification results and locate the object by bounding boxes.

YOLOv4 includes three main parts including Backbone, Neck, and Head as shown in figure 1.

- **Backbone network** will be mainly responsible for the extraction of image features.
- **Neck network** can enhance the image features and process extracted features from backbone network. Then it combines these features to improve network recognition and obtain better results.
- **Head network** applies anchor boxes on a feature map and generates a final output vector containing the class, feature, and bounding box coordinates.

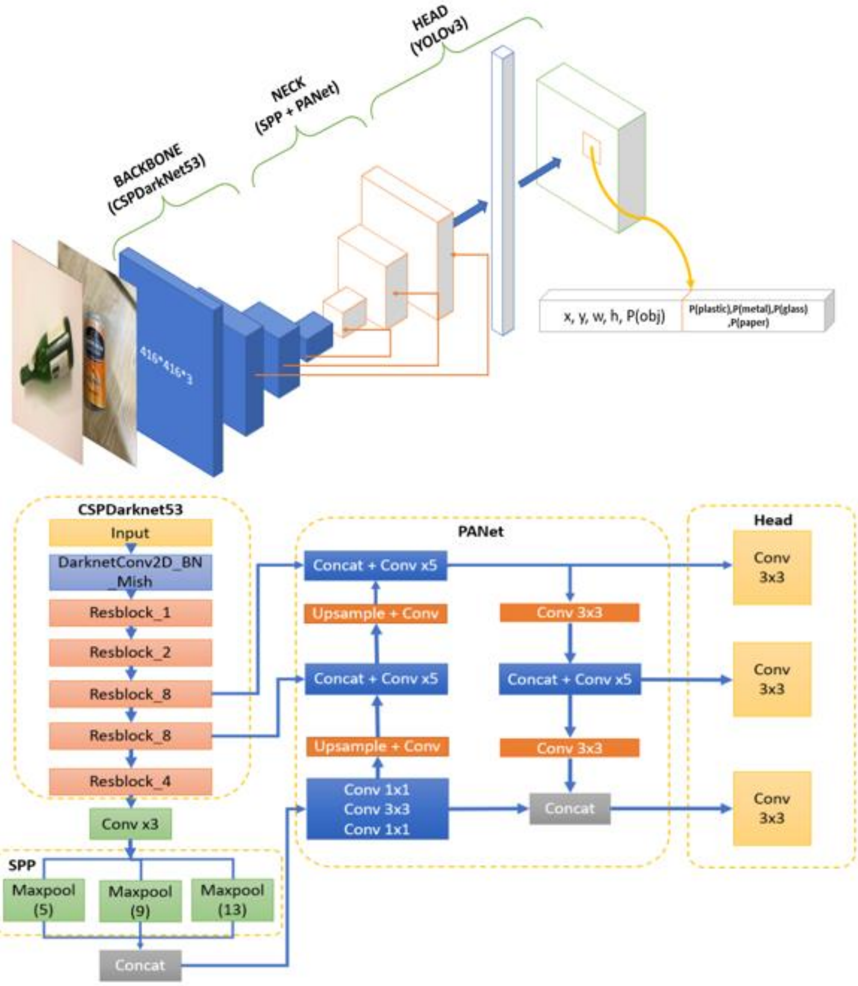

Figure 1. The overall architecture of YOLOv4 model

To calculate the loss function of YOLO will be based on three loss: Localization Loss – loss function of predicting the center coordinates, length, width of the bounding box(x, y, w, h); Classification Loss – loss function of predicting the kind of the object; Confidence Loss – loss function of predicting that the bounding box contains the object compared to the actual label. The sum formula of YOLOv4 is shown below:

$$loss = \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{obj} \left[(x_i - \overline{x}_i)^2 + \left(y_i - \overline{y}_i\right)^2 \right]$$

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 7.39 ⪤ Vol. 11, Issue 5, May 2022
DOI: 10.17148/IJARCCE.2022.11503

$$+\lambda_{coord}\sum_{i=0}^{s^2}\sum_{j=0}^{B}1_{ij}^{obj}\left[\left(\sqrt{w_i}-\sqrt{\overline{w}_i}\right)^2+\left(\sqrt{h_i}-\sqrt{\overline{h}_i}\right)^2\right]$$

$$+\sum_{i=0}^{s^2}\sum_{j=0}^{B}1_{ij}^{obj}\left(C_i-\overline{C}_i\right)^2+\lambda_{noobj}\sum_{i=0}^{s^2}\sum_{j=0}^{B}1_{ij}^{noobj}\left(C_i-\overline{C}_i\right)^2$$

$$+\sum_{i=0}^{s^2}1_{ij}^{obj}\sum_{c\in classes}\left(p_i(c)-\overline{p}_i(c)\right)^2$$

(1)

x_i, y_i is Bounding Box of truth Location.

$\bar{x}_i, \bar{y}_i$ is Bounding Box of predicted Location.

$w_{i,} h_i$ is Bounding Box of truth size.

$\overline{w}_i, \overline{h}_i$ is Bounding Box of predicted size.

C_i is the box confidence score.

$\bar{C}_i$ is the box confidence score for the predicted object

p_i is class probabilities.

$\bar{p}_i$ is class probabilities for the predicted object.

c is set of classes.

The process of training waste detection and classification model based on the YOLOv4 model illustrated as figure 2. We take as input the images. These images through the backbone network are extracted into feature maps. In neck network, object detection is improved. Then, using Dense Prediction (one-stage) and Sparse Prediction (two-stage) to locate the bounding boxes and classify waste. The weight will be saved to use in the future.

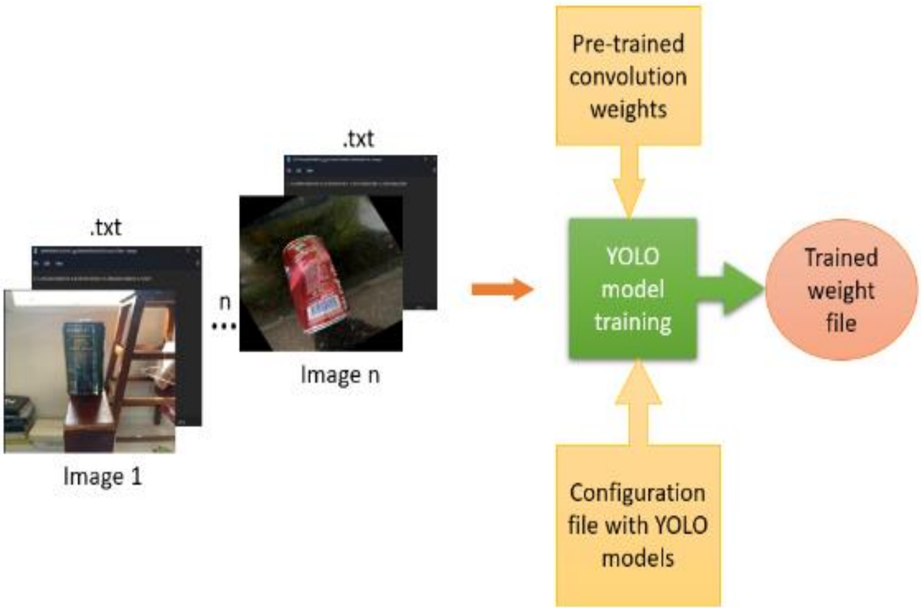

Figure 2. Model training process

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⚟ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.11503

III. EXPERIMENT

A. Dataset

We use four common kinds of waste shown in figure 3, including plastic, glass, metal, and paper.

Figure 3. Images of Waste Categories

To build a dataset for the object detection task, it is necessary to go through some stages according to a process shown in figure 4.

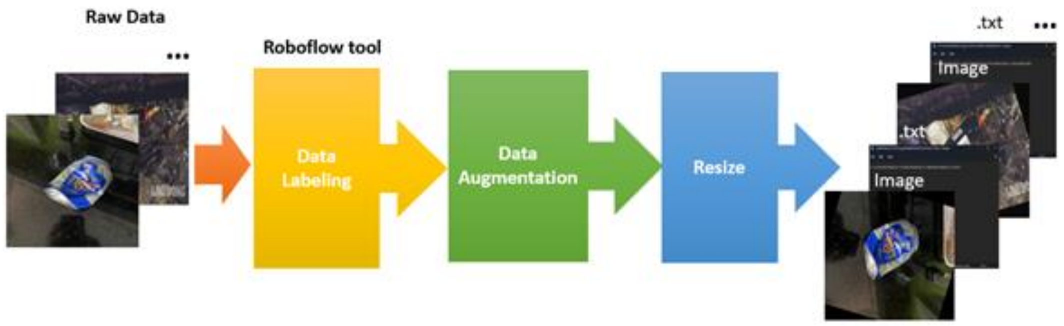

Figure 4. Data processing

1) Data collection: The dataset we built consists of 920 self-taken images and using available waste datasets such as TrashNet [8] is 500 images, TACO [9] is 750 images, and 450 images from the Drinking Waste Classification dataset [10], 388 images from Waste Pictures [11] and 49 images from the Internet, the dataset is shown in table I. Most images in the dataset have a simple background and white background, a few are actual photos such as grass background, trash can, etc.

TABLE I AMOUNT OF BOUNDING BOX

Class	Plastic	Metal	Glass	Paper	Total
Quantity	1155	922	1168	922	4237

2) Annotation data: We use the Roboflow tool to annotate the dataset with four classes from 0 to 3 respectively plastic, metal, glass, and paper. An image after annotating will be able to contain many bounding boxes and a *.txt file containing the annotation information of each object as shown in figure 5.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.11503

Figure 5. Label images within Roboflow

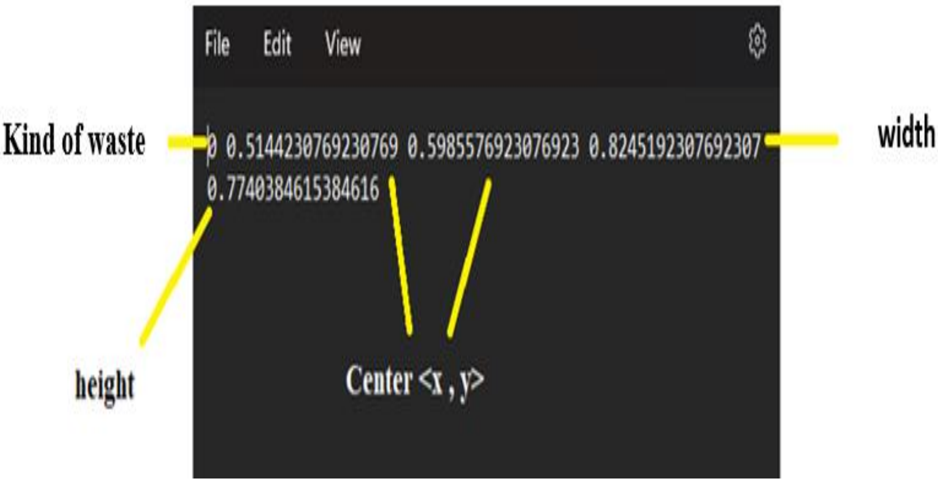

Figure 6. Bounding box annotations

Figure 6 represents the annotation information: first, the value indicates the label of an object, the next is two values are the coordinates of the center of the bounding box (x, y), and the last two values are the length and width of the bounding box measured from the center (width, height).

3) Data augmentation: we use techniques like rotating the image (-45 degrees, +45 degrees), flipping the image, and making it darker and lighter (35%). Thus, the total number of images we have includes 12228 images, Table II. These images are splitted into two subsets: 80% for training and 20% for testing.

TABLE II AMOUNT OF IMAGES IN OUR DATASET

	Raw Data	Augmentation	Total
Quantity	3057	9171	12228

B. Metric

We use a number of metrics to evaluate the results of the model such as F1-Score, AP, and mAP.

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

TP: number of true positives.

FP: number of false positives.

FN: number of false negatives

F1-Score is the harmonic mean between precision and recall.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁘ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.11503

$$F1 = 2\frac{Precision \times Recall}{Precision + Recall} \qquad (4)$$

AP is the weighted sum of precisions at each threshold where the weight is the increase in recall.

mAP is the average of AP.

$$AP = \sum_{k=0}^{k=n-1}(R_k - R_{k-1})\, P_k \qquad (5)$$

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \qquad (6)$$

R is value of Recall.
P is value of Precision.
n is number of thresholds.
AP_i is the AP of class i.
N is the number of classes.

C. Results

In this section, we present the results we have obtained. Compare the performance of the classification model when applying data augmentation and not applying data augmentation. Figure 7 shows the results of training the model, without augmenting data, we observe that the mAP of the model is about 85.36%.

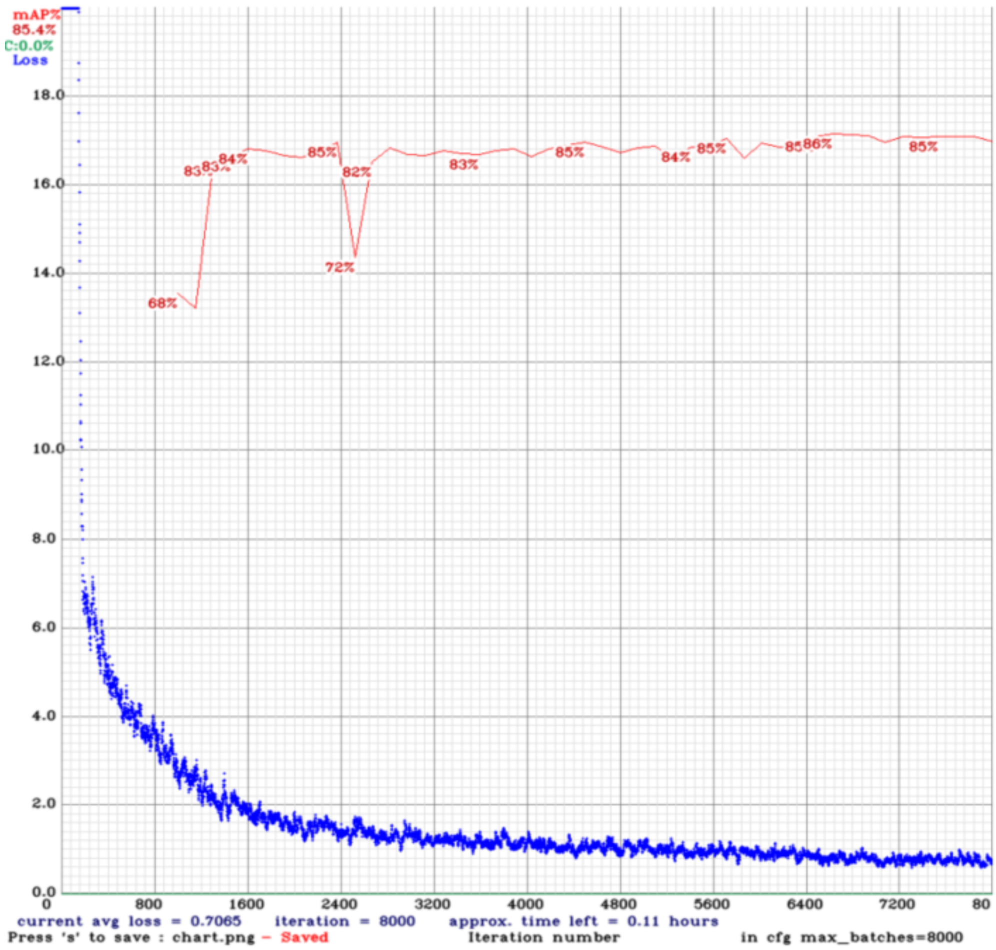

Figure 7. AP and mAP YOLOv4 in the training process without Data Augmentation

Figure 8 shows the results of the model when applying the data augmentation technique, in this way the model achieves the best performance is mAP 90.27%, we have the x-axis as the number of steps, the y-axis containing the avg loss, and the mAP%. The results of the model are shown in Table III.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⩋ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.11503

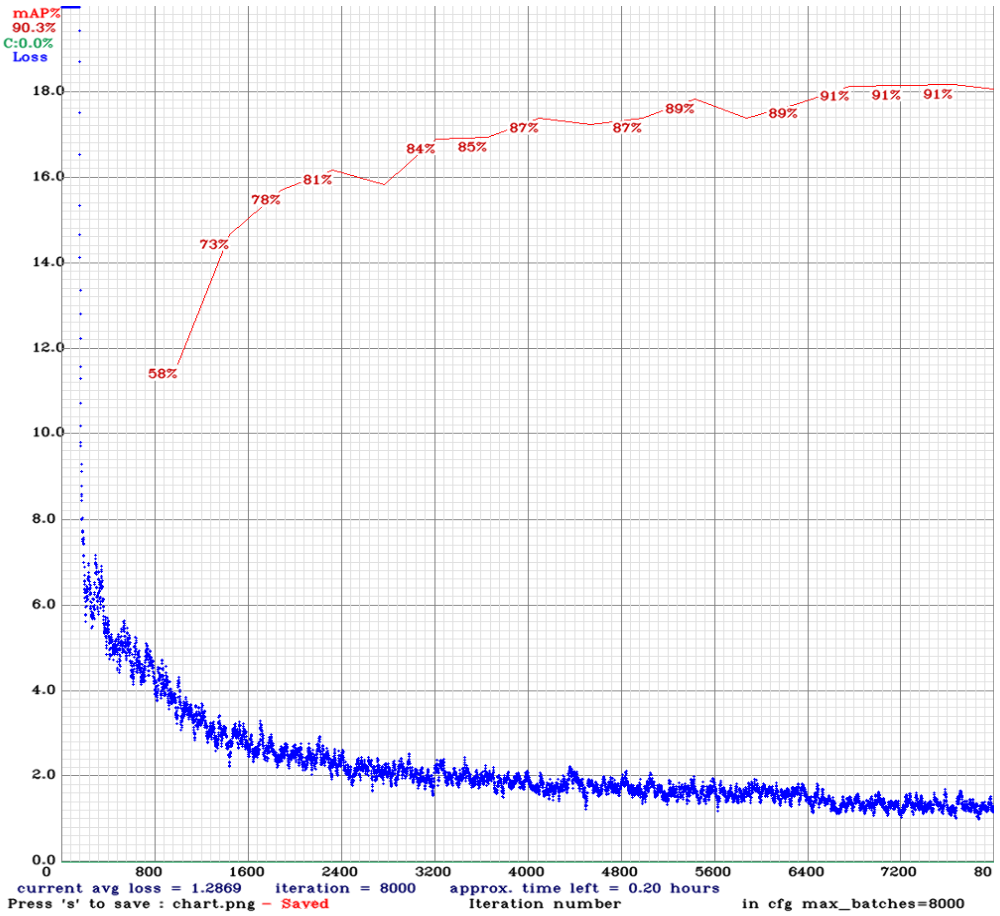

Figure 8. AP and mAP YOLOv4 in the training process with Data Augmentation

TABLE III EVALUATION MEASURE DURING TRAINING

Model	F1-score	mAP@0.50
MDA - Model with Data Augmentation	86%	90.27%
MRD - Model without Data Augmentation	83%	85.36%

TABLE IV AVERAGE PRECISION (AP)

Model	Plastic	Metal	Glass	Paper
MDA - Model with Data Augmentation	84.85%	92.96%	88.68%	94.61%
MRD - Model without Data Augmentation	72.14%	88.10%	94.24%	86.95%

Table IV, we find that the model MDA achieves promising results. The AP measurement of each layer shows that the model has the best results in the paper class 94.61%, followed by metal and glass with an accuracy of 92.96% and 88.68%, respectively, and the lowest is with plastic 84.85%.

After training the YOLOv4 model on the data that we have collected, we have achieved the requirements originally set out. In general, the model has detected and classified the object classes and is displayed as shown in figure 9, figure 10.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⋇ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.11503

Figure 9. Testing Input images using Google Colab

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁒ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.11503

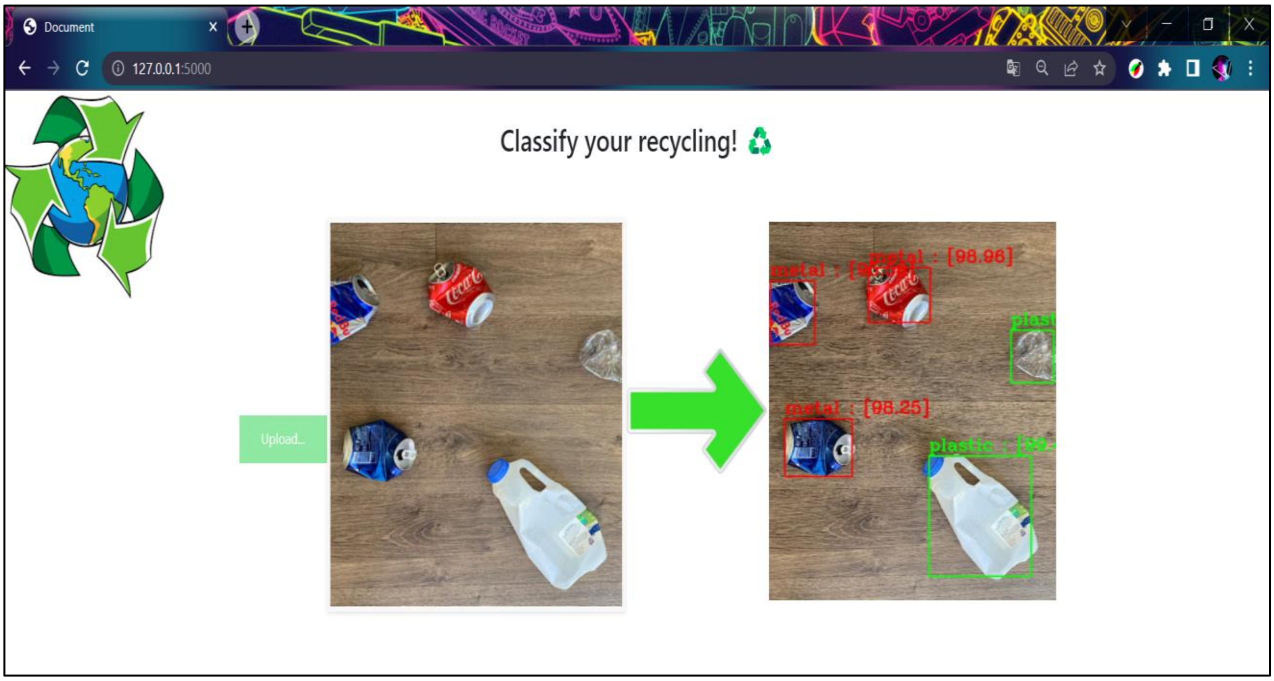

Figure 10. Web application for Waste Detection and Classification

Although there are images that are recognized correctly, there are still images where the model is not properly recognized due to the effect of the background, illumination, object properties, and the quality of the sample image. These difficulties have caused the problem to remain incompletely solved. It is shown in figure 11.

Figure 11. An example the model predicts wrong

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁘ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.11503

IV. CONCLUSION

In general, the MDA model can detect and classify well with four kinds of waste: plastic, glass, metal, and paper with an average accuracy of 84.85%, 88.68%, 92.96%, 94.61% respectively. In addition, the model has some limitations such as being easily confused with plastic bottles because of its transparent properties. From our current research results, the next development steps can be to improve and collect more data to improve the efficiency and accuracy of the model. In addition, the number of objects is also expanded, such as foam, biological waste, rubber, etc. The direction of further development in the far future is the project of automatic garbage sorting bins that can be tested in reality at the mold members of schools and public areas.

In parallel with the development, we also aim to overcome the disadvantages and problems that the model is facing such as not recognizing glass waste with lots of dirt or still confusing plastic and glass.

REFERENCES

[1] Oluwasanya Awe et al, Smart trashnet: Waste localization and classication, CS229 Project Report 2017, 2017.

[2] Meng, S., & Chu, W, A Study of Garbage Classification with Convolutional Neural Networks, 2020 Indo – Taiwan 2nd International Conference on Computing, Analytics and Networks (Indo-Taiwan ICAN), 2020, pp. 152-157.

[3] Carolis, B.D., Ladogana, F., & Macchiarulo, N. YOLO TrashNet: Garbage Detection in Video Streams. 2020 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS), 2020, pp. 1-7.

[4] Mittal, G., Yagnik, K.B., Garg, M., & Krishnan, N.C. Spotgarbage: smartphone app to detect garbage using deep learning, In Proceed-ings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, doi: 10.1145/2971648.2971731, 2016, pp. 940–945.

[5] Bochkovskiy, A., Wang, C., & Liao, H.M. YOLOv4: Optimal Speed and Accuracy of Object Detection, doi: 10.48550/arXiv.2004.10934, 2020.

[6] Juan Du, "Understanding of Object Detection Based on CNN Family and YOLO", Published under licence by IOP Publishing Ltd, doi: 10. 1088/1742-6596/1004/1/012029, 2018.

[7] Wang, S. Research Towards Yolo-Series Algorithms: Comparison and Analysis of Object Detection Models for Real-Time UAV Applications. Journal of Physics: Conference Series, 2021, pp.1948.

[8] M. Yang and G. Thung, Trashnet. (2020). Retrieved on March 2022, Available: https://github.com/garythung/trashnet.

[9] Pedro F Proença and Pedro Simões, Taco: Trash annotations in context for litter detection. doi: 10.48550/arXiv.2003.06975, 2020.

[10] Arkadiy Serezhkin, Drinking waste classification. (2020). Retrieved on March 2022, Available: https://www.kaggle.com/arkadiyhacks/drinking-waste-classification.

[11] Janusz Bobulski and Jacek Piatkowski (2018), Plastic Waste DataBase of Images – WaDaBa. Retrieved on March 2022, Available: http://wadaba.pcz.pl.