Efficient Waste Detection and Classification based on YOLOv5 Models

Suksun Promboonruang[1], Thummarat Boonrod [1] and Yongyut Ratchatawetchakul[2,*]

[1] Department of Business Computer,Faculty of Administrative Science, Kalasin University, Kalasin Sub-District, Muang District, Kalasin 46000, Thailand

[2,*] Department of Digital Business and Information System, Mahasarakham Business School,Mahasarakham University, Khamriang Sub-District, Kantarawichai District, Mahasarakham 44150, Thailand

*Corresponding Email: yongyut.r@msu.ac.th

Received November 25, 2023, Revised January 30, 2024, Accepted February 5, 2024, Published February 19, 2024

Abstract. *This paper proposes efficient waste detection and classification based on YOLOv5 by utilizing YOLOv5 for waste detection and classification. Divide the dataset into 4 classes consisting of wood, glass, plastic, and metal. The dataset is methodically divided into three subsets: the training set consisting of 1,860 images, the validation set consisting of 200 images, and the test set consisting of 235 images. The objective of our study is to assess the effectiveness of three YOLOv5 models, namely Yolov5s, Yolov5m, and Yolov5x, across several waste object categories. The methodology employed in this research is as follows: Compilation of datasets and development of models specific to each iteration of YOLOv5. Comparing models. We assess the precision, recall, and mean average precision (mAP) to measure the correctness and speed of their processing. The empirical findings from our investigation suggest that Yolov5x demonstrates the highest level of accuracy and mAP scores (0.41), whilst Yolov5s showcases the shortest processing time (0.83 hours).*

Keywords:

YOLOv5, waste detection, artificial intelligence, image processing, efficient waste detection

1. Introduction

The escalating volume of garbage has emerged as a significant environmental apprehension in Southeast Asia's swiftly evolving milieu, marked by expeditious urbanization and industrial expansion. The regional dilemma at hand exemplifies the worldwide demand for enhanced waste management techniques that are both more effective and accurate. Conventional, labor-intensive approaches are failing due to inefficiencies and inaccuracies. In this context, artificial intelligence (AI) and image processing technologies, specifically the You Only Look Once version 5 (YOLOv5) algorithm, present significant possibilities for transformation. This project aims to utilize different versions of Yolov5s, Yolov5m, and Yolov5x for innovative garbage detection and categorization, thereby making a valuable contribution to global initiatives for sustainable waste management.

The Southeast Asian region, characterized by its fast-paced urbanization and industrialization, represents a global predicament of increasing trash and that Thailand is facing global problems, which poses a challenge to current waste management methods. [5] state that traditional techniques that depend on manual sorting are susceptible to inaccuracies and inefficiencies, hence endangering the environment and human well-being. AI and image processing, namely YOLOv5, present a novel opportunity in trash management, serving as a distinct alternative. [9] showcased the capacity of artificial intelligence (AI) to improve garbage sorting processes and minimize mistakes.

The YOLO (You Only Look Once) series of object detection models has been significantly impacting various real-time applications, as demonstrated in recent research. Bandukwala et al. emphasized the precision of YOLOv4 in detecting multiple objects, boasting an accuracy of 98% for images and 99% for videos, and identified the combination of YOLOv4 with Deep SORT as optimal for vehicle counting despite occlusion and reduced visibility challenges [1]. Similarly, Chaudhari et al. presented the YOLO Real Time system, which can effectively classify, detect, and localize multiple objects in complex images, framing object detection as a regression problem and addressing the semantic gap in object detection [2]. Biju, George, and K. H. highlighted the YOLO algorithm's ability to provide stable and instantaneous object detection with high precision and speed, making it suitable for quick recognition tasks [3]. Gore, Bhasin, and S's research with YOLOv5 demonstrated high accuracy in traffic sign detection, outperforming other methods and suggesting future exploration in adverse conditions and autonomous driving systems [4]. In more specialized applications, Kiyokawa et al. developed a robotic waste sorting system with agile manipulation techniques, employing a quickly trainable detector that demonstrated high performance on items like aluminum cans, glass bottles, and plastic bottles [6]. Kusumah et al. found that the SGD optimizer outperformed others like

Adam in a YOLOv5-based acne detection model, indicating the impact of optimizer choice on model performance [7]. Yu, Liu, and Wu improved YOLO's detection of small and obscured road targets, enhancing accuracy for detecting various vehicles [8]. Tahir, Ahmad Khalid, and Mohd Fadzil used YOLOv5 to develop a child detection model that outperformed YOLOv5s in restricted areas, with potential future applications in tracking missing children [10]. Wang et al. proposed an improved YOLOv5 for smoke detection, achieving a 4.4% improvement in mAP and a detection speed of 85 FPS, indicating its suitability for engineering applications [11]. Lastly, Wani et al. trained the YOLO algorithm for vehicle crash detection with high precision and recall, enabling the communication of alerts to emergency vehicles [12], and M et al. utilized a customized YOLO framework for detecting waste and vehicles, suggesting further improvements with an expanded image database [13].

AI plays a significant role in enhancing efficiency and sustainability in garbage management. The objective of this study is to enhance the efficiency, accuracy, and sustainability of systems by investigating the capabilities and trade-offs of YOLOv5 models. This study extends the work of Wang et al. [11] and concentrates on identifying smoke by extracting distinctive characteristics, whereas Shao et al. [9] improved the identification of vehicles using an improved YOLOv5s method. The focus of our research is to utilize YOLOv5 models for sustainable waste management, addressing a notable deficiency and making a valuable contribution to the overarching objective of sustainability. This research emphasizes its significant contribution to addressing the pressing need for creative waste management solutions through careful experimentation and evaluation, utilizing metrics such as precision, recall, and mean average precision (mAP). Lack of research or knowledge gap Current research demonstrates the use of the You Only Look Once version 5 (YOLOv5) algorithm in the area of waste detection and classification. Especially for various types of waste materials. This is despite increasing evidence of the effectiveness of artificial intelligence (AI) and image processing technology in areas such as smoke detection and vehicle identification. But a detailed comparative analysis of different versions of YOLOv5, especially Yolov5s, Yolov5m, and Yolov5x, designed specifically for sustainable waste management, is lacking. This gap is extremely important. This is because it hinders the effective use of these cutting-edge technologies in improving waste management systems. This is an important requirement in areas such as Thailand in Southeast Asia. The demands of rapid urbanization and industrial growth call for effective waste management solutions. The crucial research question that arises from this gap is: in the field of waste detection and classification, how do the various versions of Yolov5s, Yolov5m, and Yolov5x perform when it comes to different types of waste materials, and which version achieves the best balance between accuracy and processing speed, which is essential for sustainable waste management? This research aims to examine and compare the capabilities and inherent trade-offs of several YOLOv5 models in the specific context of waste identification. The research aims to provide empirical insights into the performance of these models, evaluated using metrics such as precision, recall, and mean average precision (mAP).

In the field of waste separation research, there has been a notable transition from initial rudimentary item identification methods that had limited accuracy and speed in differentiating different forms of waste, to more sophisticated approaches like as YOLOv5 and YOLO-MS. These contemporary models provide enhanced precision, velocity, and the capacity to manage intricate, practical situations. The transition from fundamental research to cutting-edge, immediate solutions demonstrates the improvement of technology and a change in approach that incorporates advanced deep learning techniques, signifying a significant advancement in the field. The user's text consists of two references, [19] and [20].

2. Materials and Methods

2.1 Collection of Data

This study is based on the data from the "Waste Detection Image Dataset Version 2" obtained from Roboflow's repository. The dataset may be found at the following URL: https://universe.roboflow.com/waste-vsvfz/waste-detection-train-test/dataset/2. The dataset is methodically divided into three subsets: the training set consisting of 1,860 images, the validation set consisting of 200 images, and the test set consisting of 235 images. Every image in these collections has undergone a meticulous preparation routine, which includes automatic orientation adjustment to ensure correct alignment. In addition, to increase the strength and diversity of the dataset, augmentation techniques such as horizontal and vertical flipping, as well as 90-degree rotation (in clockwise, counter-clockwise, and upside-down directions), were utilized. Each image must indicate which class is in which position. There can be many classes, and a file format that is appropriate for the model must be created to specify what classes and positions the information in the image consists of. It offers a wide range of images that accurately represent complex detection scenarios seen in the real world, as shown in Figure 1.

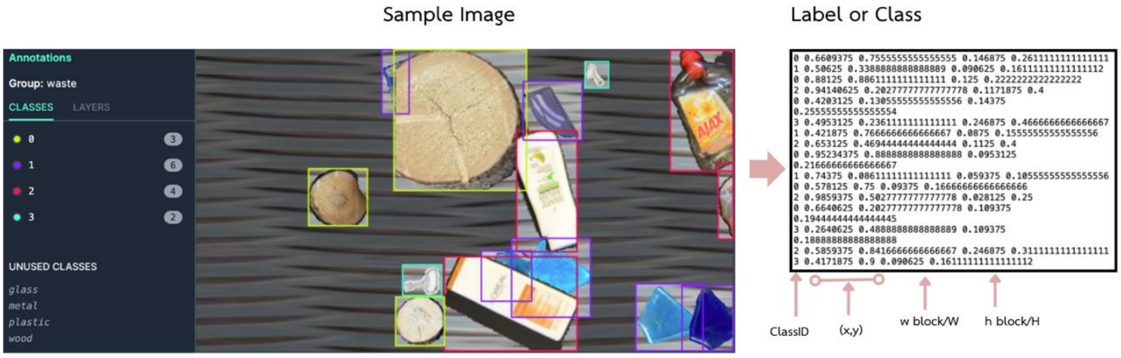

Fig. 1 Sample images from the training set.

The dataset for training and evaluating YOLOv5 models is meticulously segmented into three distinct subsets, each tailored for a specific function. The training set, comprising 1,860 images, is curated to encompass a wide spectrum of waste materials and scenarios, ensuring a comprehensive representation that augments the robustness of the models. This is followed by the validation set, which includes 200 images, playing a pivotal role in gauging the models' generalization capabilities and performance. Its diversity is instrumental in ensuring the reliability and accuracy of model assessments. Lastly, the test set, containing 235 images, acts as the ultimate measure for appraising the effectiveness and practical applicability of the models in real-world settings. The varied nature of this set is crucial for affirming the models' proficiency in accurately detecting and classifying waste in diverse environmental conditions and scenarios.

Data preprocessing

refers to the steps used to clean, transform, and prepare raw data for analysis. The dataset underwent rigorous data preprocessing to confirm its appropriateness for training and assessing YOLOv5 models. This entailed automatically adjusting the orientation of each image to ensure proper alignment, so reducing potential distortions that could impact the accuracy of the models' predictions. Furthermore, every image was carefully annotated to indicate the classes and placements of the waste materials shown, allowing the models to precisely recognize and categorize the many trash types represented in the images.

Augmentation approaches

To enhance the resilience and diversity of the dataset, a series of augmentation approaches were used. The techniques included horizontal and vertical mirroring, as well as 90-degree rotation in clockwise, counterclockwise, and inverted orientations. The intentional utilization of augmentation techniques with the objective of introducing diversity and intricacy into the dataset, replicating the various environmental circumstances and scenarios seen in actual waste management situations. By subjecting the models to augmented data, the resilience and ability to apply knowledge to various real-life scenarios of the trained YOLOv5 models were enhanced, guaranteeing their efficacy in precisely identifying and categorizing waste materials.

2.2 Methods

This Yolov5x Network Design

The YOLOv5x model, as specified in the custom_yolov5x.yaml file, embodies the highest level of complexity and capabilities within the YOLOv5 series. With a depth ratio of 1.33 and a width ratio of 1.25, it is the most profound and broadest compared to similar models. Its purpose is to effectively capture even the most intricate details for precise object detection. The anchor boxes are specifically calibrated for three scales (P3/8, P4/16, and P5/32), guaranteeing complete representation of objects of different sizes. The YOLOv5x model is constructed using convolutional (Conv) functions arranged in layers, with subsequent utilization of multiple C3 functions, which are enhanced iterations of the conventional bottleneck structure. The C3 layers play a crucial role in increasing the model's depth, enabling the extraction of subtle spatial features. The inclusion of the Spatial Pyramid Pooling Fast (SPPF) layer enhances the model's ability to include features at different scales, which is essential for recognizing objects of various sizes. The YOLOv5x utilizes convolutional layers, upsampling, and concatenation algorithms in the head section to combine feature maps from various backbone levels. The fusion process guarantees that the final detection layer gains an abundant mixture of spatial information from different resolutions. The model reaches its highest point with a detection layer, utilizing the anchors to recognize objects at different scales P3, P4, and P5. YOLOv5 is the latest installment in the series of real-time object detection systems developed by Ultralytics. Utilizing PyTorch as its main framework, YOLOv5 offers superior performance in many aspects compared to YOLOv4 and the previous version, YOLOv3, developed by Darknet [15,16] The architecture of YOLOv5 is designed to be flexible and adaptable to different problem requirements, presenting variants ranging from YOLOv5s (small) to YOLOv5x (xlarge) in size and complexity. Each variant has customized anchor boxes and network layers to enhance the accuracy of detecting various object sizes. This architecture includes multi-layer deep learning techniques such as cross-stage partial networks (CSP), spatial pyramid pooling (SPP), and PANet pathways, which collectively enable the system to detect objects quickly and accurately in a variety of environments [17]. Studies and tests have shown that YOLOv5 not only has superior processing speed but also improved detection accuracy [14]. The loss function in YOLOv5 is a comprehensive metric that aims to maximize different elements of the model's performance, including objectness, classification, and localization losses. The loss function is designed to penalize the model for making wrong predictions and is formulated to minimize the discrepancy between the predicted values and the actual values. The total lost L in YOLOv5 can be defined as:

$$\mathbf{L} = \boldsymbol{L_{obj}} + \boldsymbol{L_{cls}} + \boldsymbol{L_{box}} \qquad (1)$$

Where L_{obj} is the objectness loss, L_{cls} is the classification loss, and L_{box} is the box regression loss.

This research presents improvements to The YOLOv5x model, as specified in the custom_yolov5x.yaml file, embodies the highest level of complexity and capabilities within the YOLOv5 series. With a depth ratio of 1.33 and a width ratio of 1.25, it is the most profound and broadest compared to similar models. Its purpose is to effectively capture even the most intricate details for precise object detection. The anchor boxes are specifically calibrated for three scales (P3/8, P4/16, and P5/32), guaranteeing complete representation of objects of different sizes. The YOLOv5x model is constructed using convolutional (Conv) functions arranged in layers, with subsequent utilization of multiple C3 functions, which are enhanced iterations of the conventional bottleneck structure. The C3 layers play a crucial role in increasing the model's depth, enabling the extraction of subtle spatial features. The inclusion of the Spatial Pyramid Pooling Fast (SPPF) layer enhances the model's ability to include features at different scales, which is essential for recognizing objects of various sizes.

The YOLOv5x utilizes convolutional layers, upsampling, and concatenation algorithms in the head section to combine feature maps from various backbone levels. The fusion process guarantees that the final detection layer gains an abundant mixture of spatial information from different resolutions. The model reaches its highest point with a detection layer, utilizing the anchors to recognize objects at different scales P3, P4, and P5.

In general, the YOLOv5x configuration embodies an advanced and cutting-edge design that focuses on achieving the highest level of accuracy and level of detail in activities related to object detection. The intricate design of the architecture makes it especially suitable for challenging applications that require utmost accuracy as shown in Figure 2.

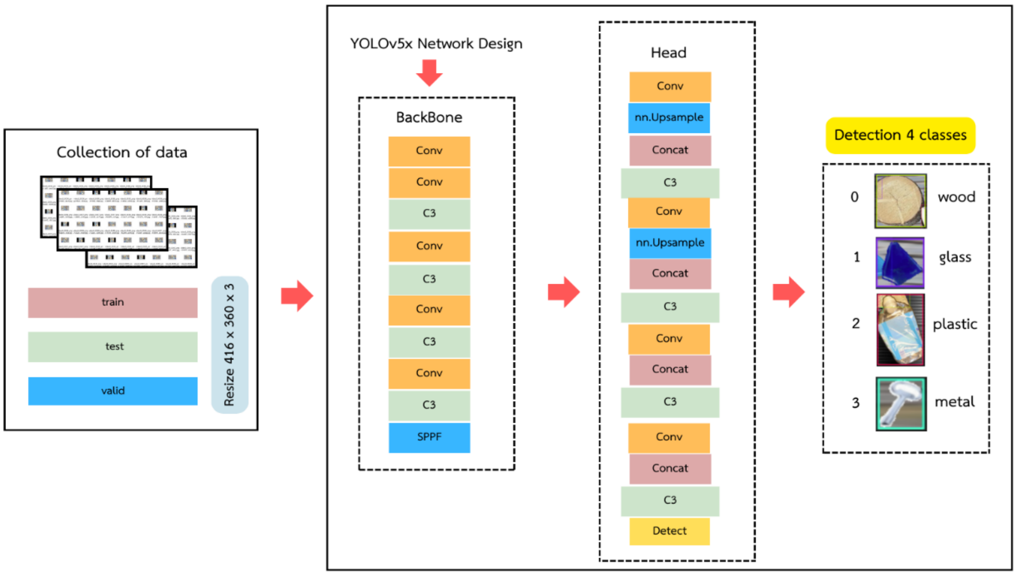

Fig. 2 Illustration of the proposed Yolov5x network design.

2.3 Model Evaluation

The assessment of this work is performed using a set of essential criteria, including precision, recall, average precision (AP), and mean average precision (mAP). These measurements are utilized to assess the precision of the models, where AP refers to the area under the precision-recall curve and mAP represents the average AP value for each category. [11] The formula is as follows:

$$\mathbf{P} = \frac{\mathbf{TP}}{\mathbf{TP+FP}} \qquad (2)$$

$$\mathbf{R} = \frac{\mathbf{TP}}{\mathbf{TP+FP}} \qquad (3)$$

$$\mathbf{AP} = \frac{\sum \mathbf{P}}{\mathbf{Num(objects)}} \qquad (4)$$

$$\mathbf{mAP} = \frac{\Sigma \mathbf{AP}}{\mathbf{Num(class)}} \quad (5)$$

where TP is the number of correct classes predicted to be correct, FN is the number of correct classes predicted to be negative, and FP is the number of negative classes predicted to be correct.

3. Experiments and Evaluations

3.1 Results

The comparative analysis of the YOLOv5 variants YOLOv5s, YOLOv5m, and YOLOv5 provides a detailed understanding of the performance differences across object detection models. This assessment thoroughly evaluates different aspects of efficacy, including accuracy, retrieval, mAP@.5, and mAP@.5:.95. The examination encompasses the identification of specific object categories, specifically glass, metal, plastic, and wood. This thorough evaluation not only highlights the skills and constraints of each model but also emphasizes the delicate equilibrium between model intricacy, training effectiveness, and detection precision.

Evaluation of YOLOv5s, YOLOv5m, and YOLOv5x Models The comparative examination of the YOLOv5 models, namely YOLOv5s, YOLOv5m, and YOLOv5x, provides a compelling narrative regarding the trade-offs between model efficiency and complexity. The YOLOv5s model is notable for its short training time of 0.830 hours and small file size of 14.4 MB. It consists of 218 layers and has a total of 7,049,425 parameters. This model requires a processing capacity of 15.7 GFLOPs. On the other hand, the YOLOv5m model has 290 layers and 20,865,057 parameters, which results in a training time of 0.893 hours and a file size of 42.1 MB. Additionally, it requires 47.9 GFLOPs. The YOLOv5x model represents a notable increase in complexity, requiring the longest training time of 1.997 hours and a huge file size of 173 MB. The model consists of 444 layers and has a total of 86,193,601 parameters, which leads to a significant computational burden of 203.8 GFLOPs. This analysis highlights the complicated relationship between training efficiency, model size, structural complexity, and processing requirements. It provides significant insights for choosing the right model depending on individual application requirements and resource limitations as shown in Table 1.

Table 1 Performance of YOLOv5s, YOLOv5m, and YOLOv5x Models.

Models	Training Time (hours)	Model Size (MB)	Layers	Parameters	GFLOPs
YOLOv5s	0.830	14.4	218	7,049,425	15.7
YOLOv5m	0.893	42.1	290	20,865,057	47.9
YOLOv5x	1.997	173	444	86,193,601	203.8

The performance measurements of the YOLOv5 versions, namely YOLOv5s, YOLOv5m, and YOLOv5x, provide detailed insights into their detection capabilities. The YOLOv5s model has a precision of 0.781 and a recall of 0.582. It achieves a mean average precision (mAP) of 0.624 at the IoU threshold of 0.5 and a mAP ranging from 0.5 to 0.95 of 0.392. The YOLOv5m model demonstrates a minor improvement in precision, achieving a score of 0.806. However, it experiences a modest decrease in recall, reaching a value of 0.546. The model also achieves mAP@.5 and mAP@.5:.95 values of 0.593 and 0.381, respectively. The YOLOv5x model shows similar precision to the YOLOv5s model at 0.783, but it has a balanced recall of 0.564 and performs better in the comprehensive mAP metrics with 0.617 at IoU 0.5 and 0.417 across a range of IoU thresholds from 0.5 to 0.95. This research emphasizes the complex trade-offs between precision and recall among the models and emphasizes the significance of including the mAP scores for a comprehensive comprehension of model performance in object detection tasks as shown in Figure 3.

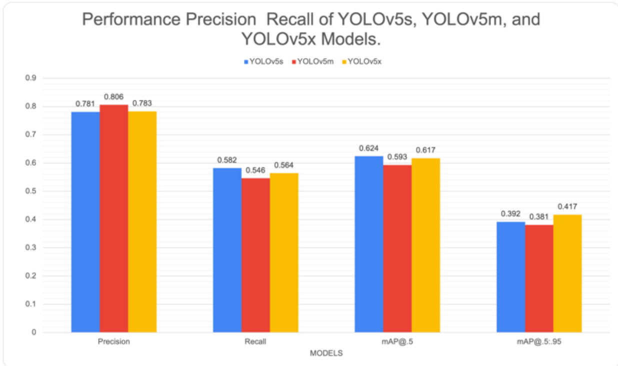

Fig. 3 Performance Precision Recall of YOLOv5s, YOLOv5m, and YOLOv5x Models.

3.2 Model Performance by Object Type

The YOLOv5 models, namely YOLOv5s, YOLOv5m, and YOLOv5x, exhibit diverse detection capabilities when evaluated using object-specific performance metrics. The YOLOv5s model has exceptional proficiency in detecting glass, with a precision of 0.862. This indicates its robust capability to accurately recognize glass items. The YOLOv5m model has exceptional accuracy in identifying metal items, with a score of 0.879. The YOLOv5x model excels at plastic detection, demonstrating an impressive precision of 0.911. Nevertheless, the YOLOv5s model shows a decrease in performance when identifying wood, with a precision of 0.647. The recall rates vary across the models, indicating differences in their capacity to accurately identify all pertinent objects within the categories. The mAP scores, calculated at IoU thresholds of 0.5 and ranging from 0.5 to 0.95, highlight the abilities and constraints of each model in accurately and comprehensively detecting various sorts of objects. This investigation highlights the importance of taking into account both accuracy and recall, as well as mAP scores, when evaluating the effectiveness of a model in various object identification scenarios as shown in Table 3.

Table 3 Model Performance by Object Type.

Object	Model	Precision	Recall	mAP@.5	mAP@.5:.95
Glass	YOLOv5s	0.862	0.591	0.674	0.446
	YOLOv5m	0.624	0.238	0.259	0.114
	YOLOv5x	0.791	0.729	0.753	0.484
Metal	YOLOv5s	0.858	0.775	0.815	0.561
	YOLOv5m	0.879	0.571	0.658	0.439
	YOLOv5x	0.517	0.221	0.242	0.107
Plastic	YOLOv5s	0.757	0.756	0.764	0.464
	YOLOv5m	0.848	0.686	0.734	0.528
	YOLOv5x	0.911	0.587	0.680	0.478
Wood	YOLOv5s	0.647	0.205	0.242	0.0958
	YOLOv5m	0.873	0.687	0.723	0.443
	YOLOv5x	0.913	0.721	0.793	0.599

The YOLOv5 models have undergone thorough evaluation to assess their varied capabilities in the field of object detection. The YOLOv5s model is notable for its fast training and compact file size. It demonstrates proficiency in distinguishing wooden and plastic objects, obtaining impressive precision and recall metrics. In contrast, the YOLOv5m model, which has a larger file size and slightly longer training time than its counterpart, has exceptional performance in accurately detecting glass and metal items. Finally, the YOLOv5x model, although requiring the longest training time and having significant computational complexity, stands out as the epitome of adaptability. It exhibits exceptional ability in recognizing various sorts of objects, particularly in identifying glass and wood. This thorough examination highlights the subtle trade-offs among model efficiency, complexity, and detection accuracy in the YOLOv5 suite. Especially suitable for challenging applications that require utmost accuracy as shown in Figure 3 and Figure 4.

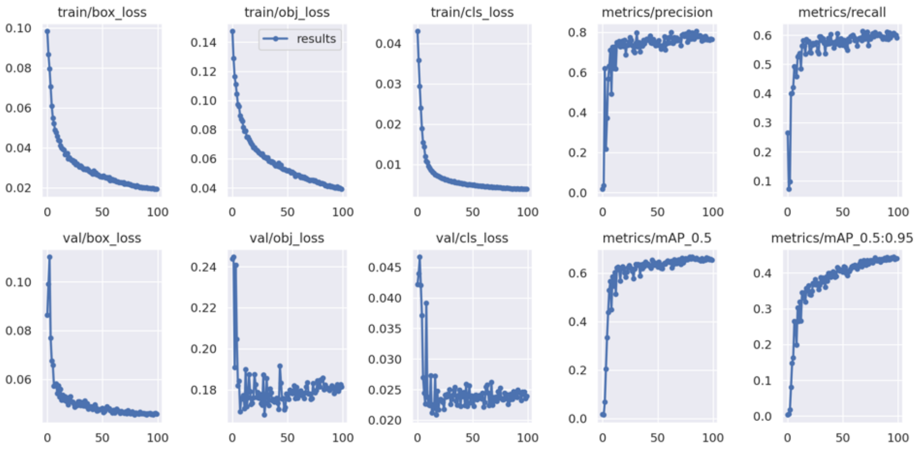

Fig. 3 The training results of YOLOv5x.

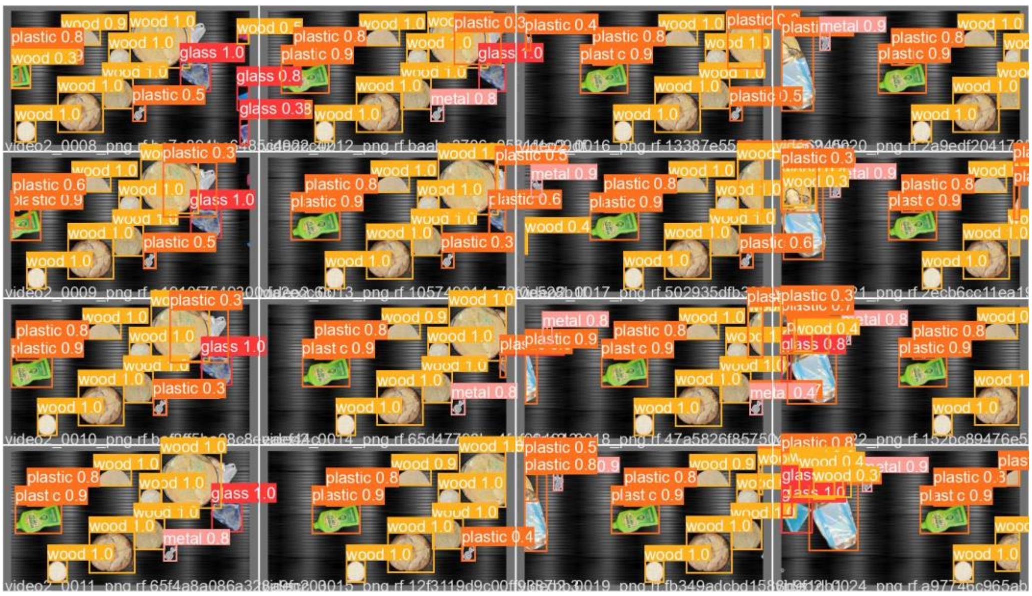

Fig. 4 Detection best results of YOLOv5x.

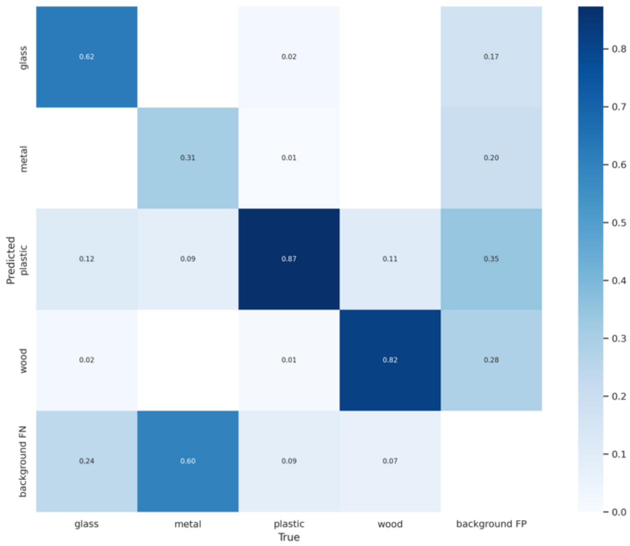

Fig. 5 Detection best confusion matrix of YOLOv5x.

4. Conclusions

The examination of the several variations of YOLOv5 provides a detailed understanding of how the models are specialized and optimized. Each iteration of the YOLOv5 series, including s, m, and x, demonstrates unique capabilities in the field of object detection, addressing various needs. The YOLOv5x is particularly noteworthy due to its lengthy training process and increased complexity. It distinguishes itself as the most resilient model, demonstrating outstanding effectiveness across a wide range of object categories. This discovery emphasizes the crucial importance of model selection, which requires a careful equilibrium between multiple essential aspects. These factors include the length of the training process, the size of the model's file, its intrinsic intricacy, and the level of precision needed to detect specific sorts of objects. The complex decision-making process highlights the various factors that support the use of object detection models in different applications. However, there is an important dearth of a thorough comparative examination of different versions of YOLOv5, especially YOLOv5s, YOLOv5m, and YOLOv5x, expressly tailored for the purpose of sustainable waste management. This study offers empirical observations on the effectiveness of YOLOv5 models, assessed by measures like precision, recall, and mean average precision (mAP). The comparative analysis of YOLOv5s, YOLOv5m, and YOLOv5x highlights the balance between model efficiency and complexity, emphasizing the complex relationship between model complexity, training effectiveness, and detection precision.

In this research, the focus lies not in the development of novel technology but rather in the in-depth analysis of the YOLOv5s, YOLOv5m, and YOLOv5x algorithms. The results obtained from the experimental process reveal distinct proficiencies of each model in detecting specific waste materials, shedding light on their specialized attributes and processes The exceptional proficiency of the YOLOv5s model in detecting glass, as indicated by a precision of 0.862, underscores its robust capability to accurately recognize glass items. This proficiency can be attributed to the model's specialized processes or attributes tailored to the detection of glass. Further research into the inner workings of the YOLOv5s algorithm may reveal specific features or training methodologies that enhance its ability to detect glass items with high precision. Similarly, the exceptional accuracy of the YOLOv5m model in identifying metal items, with a score of 0.879, suggests the presence of specialized processes or attributes geared towards metal detection within the model. Delving into the intricacies of the YOLOv5m algorithm may unveil specific mechanisms or training techniques that contribute to its remarkable accuracy in detecting metal items. Furthermore, the outstanding performance of the YOLOv5x model in plastic detection, demonstrating an impressive precision of 0.911, highlights its specialized attributes or processes tailored to the detection of plastic. Exploring the unique characteristics or training methodologies embedded within the YOLOv5x algorithm may elucidate the factors contributing to its exceptional precision in detecting plastic items. In summary, the distinct proficiencies of the YOLOv5s, YOLOv5m, and YOLOv5x models in detecting specific waste materials warrant further investigation into their specialized processes and attributes. By delving into the inner workings of each algorithm, a deeper understanding of their capabilities in detecting glass, metal, and plastic items can be attained, contributing to the comprehensive analysis of the experimental results.

Recommendations for Future Work should prioritize optimizing YOLOv5 models for waste detection by investigating advanced techniques to enhance accuracy and reduce computational requirements. Additionally, it is important to explore the practical application of YOLOv5 models in real-world waste management systems, evaluating their performance in various and changing environments.

Furthermore, integrating YOLOv5 models with other AI and machine learning techniques should be explored to improve waste detection and classification capabilities. Enhancing and employing more extensive datasets that encompass a broader range of waste materials and environmental variables would enhance the resilience of the model.

Acknowledgements

The authors extend their profound gratitude to Kalasin University, the Faculty of Administrative Sciences, their contribution has been instrumental in facilitating this research, providing an environment conducive to academic inquiry and innovation.

References

[1] D. Bandukwala, M. Momin, A. Khan and L. Islam, "Object Detection using YOLO," *International Journal for Research in Applied Science and Engineering Technology*, vol. 10, no. 5, pp. 823–29, DOI: 10.22214/ijraset.2022.42088, 2022.

[2] S. Chaudhari, N. Malkan, A. Momin and M. Bonde, "Yolo Real Time Object Detection," *International Journal for Research in Applied Science and Engineering Technology*, vol. 10, no. 5, pp. 823–29, DOI: 10.14445/22312803/IJCTT-V68I6P112, 2020.

[3] A. George, "Object Detection Using YOLO Algorithm," *Kristu Jayanti Journal of Computational Sciences (KJCS), Kristu Jayanti College (Autonomous)*, pp. 25–37, DOI: 10.59176/kjcs.v2i1.2219, 2022.

[4] S. Kale, D. Gore and S. Suchitra, "Traffic Sign Detection using Yolo v5," *International Journal for Research in Applied Science and Engineering Technology*, vol. 11, no. 5, pp. 2679–83, DOI: 10.22214/ijraset.2023.51591, 2023.

[5] P. Kim, Z. Fang and X. Huang, "SSD PCB Component Detection Using YOLOv5 Model," *Journal of Information and Communication Convergence Engineering*, vol. 21, no. 1, pp. 24–31, DOI: 10.56977/jicce.2023.21.1.24, 2023.

[6] T. Kiyokawa, H. Katayama, Y. Tatsuta, J. Takamatsu and T. Ogasawara, "Robotic Waste Sorter with Agile Manipulation and Quickly Trainable Detector," *IEEE Access*, vol. 9, 2021, pp. 124616–31, DOI: 10.1109/ACCESS.2021.3110795, 2021.

[7] H. Kusumah, M. S. Zahran, K. N. Rifqi, D. A. Putri and E. M. Hapsari, W, "Deep Learning Pada Detektor Jerawat: Model YOLOv5," *Journal Sensi*, vol. 9, no. 1, pp. 24–35, DOI: 10.33050/sensi.v9i1.2620 2023.

[8] H. Liu and Q. Wu, "An Improved YOLO for Road and Vehicle Target Detection Model," *Journal of ICT Standardization*, DOI: 10.13052/jicts2245-800x.1125, 2023.

[9] L. Shao, H. Wu, C. Li and J. Li, "A Vehicle Recognition Model Based on Improved YOLOv5," *Electronics*, vol. 12, no. 6, p. 1323, DOI: 10.3390/electronics12061323, 2023.

[10] A. Tahir, S. K. Khalid and L. Fadzil, "Child Detection Model Using YOLOv5," *Journal of Soft Computing and Data Mining*, vol. 4, no. 1, DOI: 10.30880/jscdm.2023.04.01.007 2023.

[11] Z. Wang, L. Wu, T. Li and P. Shi, "A Smoke Detection Model Based on Improved YOLOv5," *Mathematics*, vol. 10, no. 7, p. 1190. DOI: 10.3390/math10071190, 2022.

[12] L. K. Wani, M. M. Momin, S. Bhosale, A. Yadav and M. Nili, "Vehicle Crash Detection using YOLO Algorithm," *International Journal of Computer Science and Mobile Computing*, vol. 11, no. 5, pp. 75–82, DOI: 10.47760/ijcsmc.2022.v11i05.007, 2022.

[13] R. M, R. Y A, R. B. S, P. K. R and S. N, "Waste and Vehicle Detection Using YOLO," *International Journal for Research in Applied Science and Engineering Technology*, vol. 10, no. 5, pp. 3587–90, DOI: 10.22214/ijraset.2022.43124, 2022.

[14] G. Jocher, A. Stoken, J. Borovec, NanoCode012, L. Changyu, A. Hogan and A Rituerto, Ultralytics/yolov5: v3.0, GitHub, [Online], Available: https://github.com/ultralytics/yolov5, 2020.

[15] A. Bochkovskiy, C. Y. Wang and H. Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," arXiv preprint arXiv:2004.10934, 2020.

[16] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," arXiv preprint arXiv:1804.02767, 2018.

[17] C. Y. Wang, A. Bochkovskiy and H. Y. M. Liao, "Scaled-YOLOv4: Scaling Cross Stage Partial Network," arXiv preprint arXiv:2011.08036, 2021.

[18] I. Kumpala, N. Wichapha and P. Prasomsab, "Sugar Cane Red Stripe Disease Detection using YOLO CNN of Deep Learning Technique," Engineering Access, 8(2), 192–197, 2022, doi: 10.14456/mijet.2022.25.

[19] Y. Chen, "YOLO-MS: Rethinking Multi-Scale Representation Learning for Real-time Object Detection," arXiv:2308.05480, 2023.

[20] J. Khoramdel, "YOLO-Former: YOLO Shakes Hand With ViT," arXiv:2401.06244, 2024.