

©2025 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Vision-based Household Waste Classification and Smart Bin Development

Tien Sang Ngo¹, Thi Thu Phuong Nguyen¹, Vu Quang Nguyen¹, Tien Dung Cao¹, Dieu Truc Ha¹,
Tran Hiep Dinh^{1,*} and Quang Ha²

Corresponding author: Tran Hiep Dinh (tranhiep.dinh@vnu.edu.vn)

¹Faculty of Engineering Mechanics and Automation, VNU University of Engineering and Technology, Vietnam

²Faculty of Engineering and Information Technology, University of Technology Sydney, Australia

Abstract—The environment was negatively impacted because of the incorrect disposal of domestic waste. At the moment, the five most popular disposal methods are composting, waste compaction, landfilling, incineration, and biogas generation. In particular, landfills and incinerators are highly common in many countries; however, they may release certain pollutant gases contributing to the green house effect while there may be possibilities of waste generation. While classifying waste items is the key to waste management, household trash is usually classified incorrectly due to residents' negligence of the importance of waste classification, imperfect infrastructure, or the need for more instructions from local authorities. To address this problem, we propose a vision-based approach for waste classification and develop a prototype to test the system's real-time function. We applied transfer learning to fine-tune four well-known pre-trained object detection models on a self-collected dataset of 1100 samples. A smart trash bin (STB) prototype is developed for system validation and verification. The models are then implemented on the developed STB, and their CPU and GPU performance is tested. Experiment results have emphasized the effectiveness and feasibility of the proposed approach to address ecological problems in cities.

I. INTRODUCTION

Human life is becoming more and more modern, and science and technology are advancing quickly. However, domestic waste management is still a big problem in society. The amount of rubbish discharged into the environment has increased rapidly to dizziness. According to the World Bank [1], the amount of municipal solid trash produced worldwide is 2.01 billion tonnes per year, and global waste is expected to grow to 3.40 billion tonnes by 2050. Some of the waste disposal methods are composting, waste compaction, landfilling, incineration, biogas generation, and other techniques. Landfills and incinerators are two popular methods in many countries. Landfilling can pollute groundwater sources, and it can also emit some gases that affect human health, which is described in [2]. Food contamination is one of the effects of incinerator trash, and consuming pollutants poses a serious risk to local and remote inhabitants alike. It was also associated with neoplasia, problems with reproduction, and other illnesses [3].

The World Code for the Protection of the Environment has become more aware of environmental challenges as a result of everything that has been discussed above. One of the

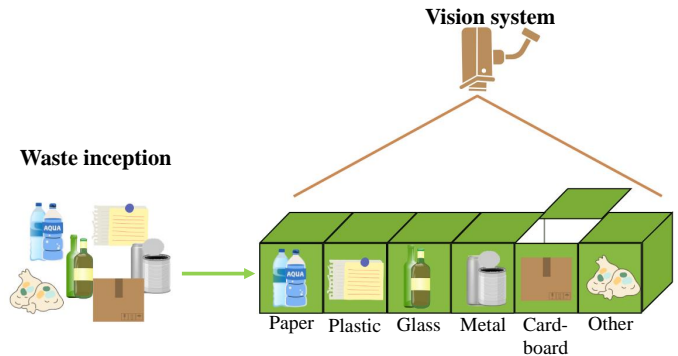


Fig. 1: A vision-based waste classification concept.

solutions that can solve this problem is to improve awareness of individuals to classify household waste. Furthermore, it can help to lower the amount of waste that enters the environment. However, the majority of waste classification techniques rely on human judgment, which can be challenging and time-consuming at times. Besides, residents' negligence of the importance of waste classification, imperfect infrastructure, or the need for detailed instructions from local authorities are also key factors contributing to the problem [4], [5]. Trash can become hazardous or non-hazardous with additional subdivisions in a different setting [6]. Specific features taken into account when classifying waste include its physical properties, the potential for biodegradation, potential for reuse, place of production, and effects on the environment [7], [8].

Numerous vision-based studies on solid waste management have been carried out to tackle this challenge, thanks to the recent advance in deep learning and edge computing [9]. The general concept is to use a camera to monitor the waste inception, the images of which are then fed into a processing board for object detection. A mechanism receives the information and then classifies the object into different categories. Fig. 1 visualizes the idea of using computer vision for household garbage classification. This system enables the automatic classification and management of waste from its initial inception through to the final sorting process. The intelligent household bin in [10] can sort four different categories of waste: hazardous waste, other waste, recyclable waste, and kitchen waste with an accuracy of 93.38% on the test set.

The proposed method [6] for domestic trash categorization with transfer learning and VGG16 achieved an accuracy of more than 96%; however, it was not implemented on edge devices. A garbage classification system based on improved ShuffleNetv2 in [11] was used to categorize recycled waste on Raspberry Pi 4, but it is not suitable for sorting domestic garbage in the house. For a separate collection of recyclables, the trash cans [12] used a tiny model of YOLO version 4. When implemented on a Raspberry Pi 4B, this model obtained an accuracy of 75%. The SMARTSORT [13] used YOLO version 4 based on Raspberry Pi 4 with 98% accuracy in their self-collected data set. To classify domestic waste, two garbage classification algorithms—YOLO version 5 and YOLO version 8 respectively in [14], [15]—were used, but neither one was used in edge devices. The deep learning models and edge devices were considered in conjunction with all of the previously mentioned elements to produce a practical smart trash bin. This study discusses two main tasks of vision-based waste management, including object classification and automatic waste management. Here, we implement some well-known classification models on our self-collected data and then develop a Smart Trash Bin (STB) to verify their performance in waste management. In the former, VGG16 [6], MobileNetV2 [16], EfficientNetV2 [17], and the newest version of the YOLO family (YOLOv10 [18]) were employed. The latter consists of the STB including a frame, a two-bin classifier, a Jetson Nano, a Camera IMX219-77, an Arduino Uno, a 16-channel servo driver, three HC-SR04 ultrasonic sensors, and eight servo motors. The contributions of our work are threefold:

- implementation and fine-tuning of some well-known classifiers on our self-collected dataset,
- development of the STB for automatic waste management, and
- comprehensive performance evaluation of the selected models and the developed prototype.

The rest of the paper is organized as follows. Section II presents the proposed methodology. Section III reports the results of the distance measurement and model evaluation. Sections IV and V discuss the findings and conclude the work.

II. METHODOLOGY

In this work, some recent effective deep learning models, i.e. VGG16, MobileNetV2, EfficientNetV2, and YOLOv10, are employed to validate the feasibility of vision-based waste classification. The VGG16, MobileNetV2, and EfficientNetV2 models were initially pre-trained on the ImageNet dataset, whereas the YOLOv10 model was pre-trained on the COCO dataset. Subsequently, the four pre-trained models are fine-tuned on our self-collected dataset. In addition to the models mentioned above, a prototype of the STB is also developed to assess the system’s effectiveness. The detection accuracy and inference time of the participating models will be evaluated on our collected data and on the Jetson Nano mounted on the STB. Algorithm 1 demonstrates the evaluation of the participating models via the classification accuracy F_1 .

Algorithm 1: Model evaluation

Input: $M = [M_1, M_2, M_3, M_4]$; // Pre-trained models
Input: $I = [I_1, I_2, \dots, I_N]$; // Our collected images
Output: M_{FT}^* ; // Best model with highest F_1

- 1 Initialize $M_{FT} \leftarrow []$;
- 2 **for** $i = 1$ **to** 4 **do**
- 3 $M_{FT} \leftarrow \text{Fine_tune_model}(M[i])$;
- 4 $M_{FT}.\text{append}(M_{FT})$;
- 5 Initialize $F1_scores \leftarrow []$;
- 6 **for** $m = 1$ **to** 4 **do**
- 7 Initialize $model_f1 \leftarrow []$;
- 8 **for** $j = 1$ **to** N **do**
- 9 $f1 \leftarrow \text{evaluate_model}(M_{FT}[m], I[j])$;
- 10 $model_f1.\text{append}(f1)$;
- 11 $avg_f1 \leftarrow \frac{\sum model_f1}{N}$;
- 12 $F1_scores.\text{append}(avg_f1)$;
- 13 $max_score \leftarrow \max(F1_scores)$;
- 14 $best_model_index \leftarrow F1_scores.\text{index}(max_score)$;
- 15 $M_{FT}^* \leftarrow M_{FT}[best_model_index]$;
- 16 **return** M_{FT}^* ;

The design of the smart trash bin is briefed in subsection A. Subsection B details an explanation of the hardware components and a wiring diagram. Subsection C describes the prototype of the smart trash bin. Section D explains the HC-SR04 sensor experiment, including its purpose and methods. The deep learning models are introduced in Subsection E. Subsection F includes self-collection data and metrics evaluation.

A. Design of the STB

As shown in Fig. 2, the smart trash bin is divided physically into two main parts: a frame and a two-bin classifier.

The frame: comprises two modules: one for the classification task and the other for the control of the waste management. The camera takes photos of the waste, which then be processed by the classifier embedded in the Jetson Nano. Depending on the classification results, the Arduino board will receive signals A or B from Jetson Nano. The waste will be sorted into bins A or B by a 16-channel servo control that is handled by the Arduino of the main control module.

The two-bin classifier: has three primary purposes. First, the distance to the user obtained by the ultrasonic sensor allows the door to open and close automatically. This mechanism reduces direct contact and stops bacterial transmission. Upon receiving a signal of the incoming waste, the automatic classifier processed by the Jetson Nano triggers the opening of the corresponding lid. Eventually, the fill level of each bin is indicated by the distances measured by the other two ultrasonic sensors, avoiding the overflowed waste.

The processing pipeline of the smart trash bin is shown in Fig. 3, Utra 1, 2, and 3 refer to the ultrasonic sensor HC-SR04 measuring the distance to the user and fill level of bins A and B, respectively. D1, D2, and D3 stand for the corresponding measured distances of Utra 1,2,3 in centimetres.

In the first step, D2 and D3 indicate the fullness of bins A and B. If D2 or D3 is larger than a threshold T_{bin} , the empty level is met, and the measured distance D1 is then considered. If D1 is less than a threshold T_{user} , the door will be opened. In

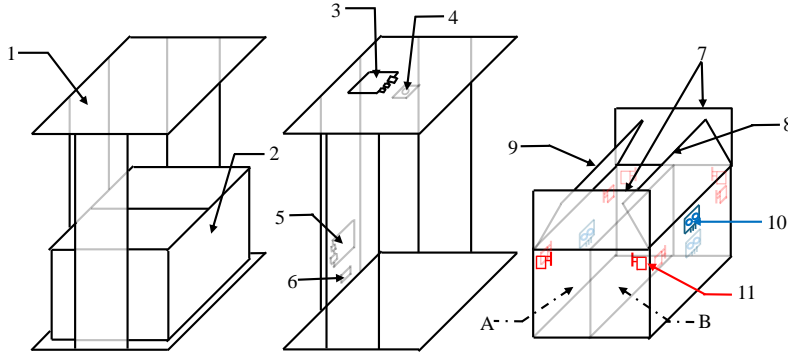


Fig. 2: Overall design of the STB. 1: Frame, 2: Two-bin classifier, 3: Jetson nano, 4: Camera IMX219-77, 5: Arduino Uno. 6: 16-channel servo control, 7: Door cover, 8: Lid cover A, 9: Lid cover B, 10: HC-SR04 ultrasonic sensors, 11: Servo motors.

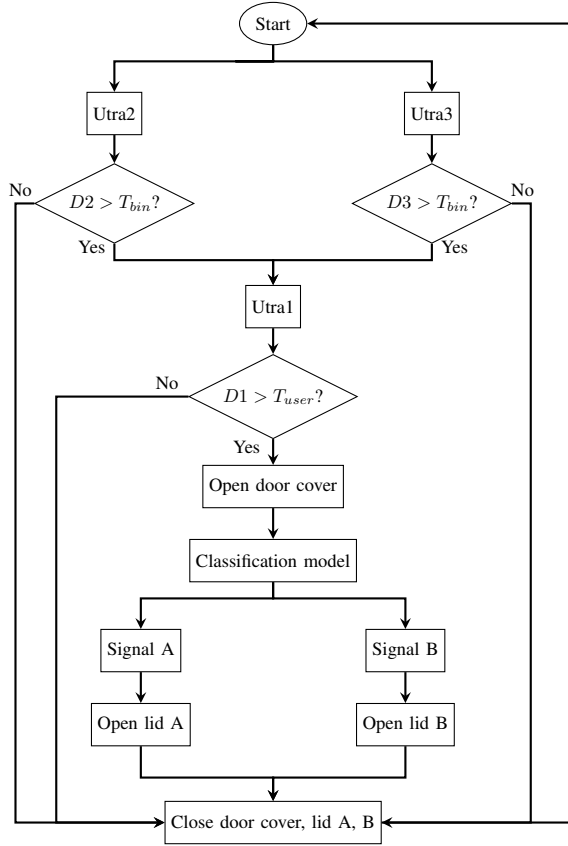


Fig. 3: Processing pipeline of the smart trash bin

this prototype, T_{bin} and T_{user} are respectively set as 3 and 10 cm. In the final step, when a classification signal is sent from the Jetson Nano to the Arduino by the classification model, bin A is opened accordingly if the signal is A and vice versa.

B. Hardware and wiring diagram

The wiring diagram is described in Fig. 4. The Developer Kit employed is NVIDIA Jetson Nano B01 with 4GB RAM. The IMX219-77 camera and the Jetson Nano board are connected to carry out the object classification task. An A-B cable is used to transfer the classification signal from the Jetson

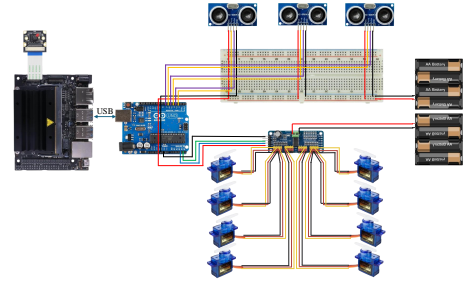


Fig. 4: Overall circuit connection

Nano to the Arduino board. The Arduino board serves as the primary controller in this project. The pins 2, 3, 4, 5, 6, and 7 of the board are used to link to the three HC-SR04 ultrasonic sensors. The pin pairings that connect the 16-channel servo controller to the Arduino are VCC-5V, GND-GND, SDA-A4, and SCL-A5. By means of matching pins, the 16-channel servo controller is linked to the eight servo motors.

The developed prototype of the STB is shown in Fig. 5, consisting of the mentioned hardware equipped on the wooden frame and the plastic two-bin classifier. The current version allows an intelligent management of two types of waste: bottles and cans, which are classified by a fine-tuned pre-trained model.

C. Estimation of measurement number with HC-SR04

To obtain a close-to-expected value of the distances to be measured, data obtained by the ultrasonic sensors are averaged for further processing at the Arduino board. Although a sufficiently large number of measurements n would return a better estimation of the true value, high n also leads to the delay in processing. As suggested in [19] the experimental standard deviation of the mean (SDM) is selected in this work for the determination of n . Let \bar{q} be the arithmetic mean of n consecutive measurements, q_k be the k th obtained value of the measurands, the standard deviation (SD) of the measured

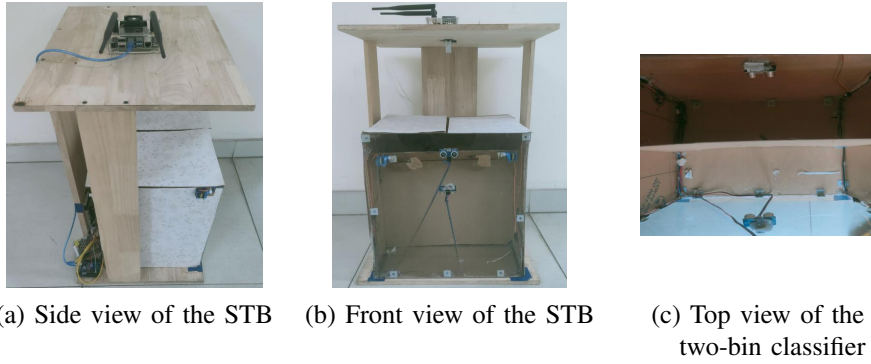


Fig. 5: The STB prototype

values $s(q_k)$ is calculated as

$$s(q_k) = \sqrt{\frac{\sum_{j=1}^n (q_j - \bar{q})^2}{n-1}}. \quad (1)$$

The SDM $s(\bar{q})$ is given as

$$s(\bar{q}) = \frac{s(q_k)}{\sqrt{n}}. \quad (2)$$

The number of measurements n is selected when the difference between the SDM calculated at two consecutive observations does not exceed a pre-defined threshold Δ , i.e.

$$s^*(\bar{q}) = \{s(\bar{q}) \mid \frac{s(q_k) - s(q_{k+1})}{s(q_k)} < \Delta\}. \quad (3)$$

In this work, the threshold is selected as 5% of the analyzed standard deviation of the mean $s(\bar{q})$.

D. Learning models

The participating models in this work are VGG16 [6], MobileNetV2 [16], EfficientNetV2 [17], and the recently published YOLOv10 [18]. The details of the models are introduced below.

VGG16 is a convolutional neural network model for image recognition. It's unique because it has only 16 layers with weights instead of relying on many hyper-parameters. It's considered one of the best vision model architectures.

MobileNetV2 is a lightweight convolutional neural network and it is specifically designed for mobile and embedded vision applications

EfficientNetV2 is also a lightweight convolutional neural network architecture and scaling technique that uses a compound coefficient to scale all dimensions of depth, width, and resolution equally.

YOLOv10 Only Look Once (YOLO) proposes using an end-to-end neural network that makes predictions of bounding boxes and class probabilities all at once.

The first three models were trained on the ImageNet dataset [20] while the last one is trained on the COCO dataset [21].

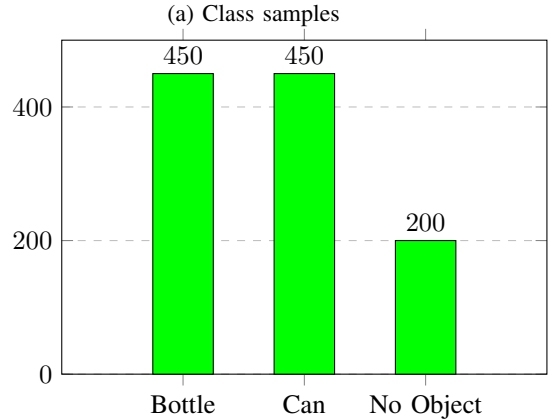


Fig. 6: Evaluation dataset.

E. Data and evaluation metrics

The proposed approach is evaluated on our self-collected data, consisting of 1100 images of size 1920×1080 and split with a ratio of 8:1:1 for training, validation, and testing. The data includes three classes, i.e. bottle, can, and no object, the sample sizes of which are 450, 450, and 200, respectively. Fig. 6 shows class samples and class distribution. Besides the F_1 score, the performance of the participating models is also evaluated via the following metrics: accuracy (Acc), precision (Pr), recall (Rc), and Frames per Second (Fps).

III. RESULTS

A. Distance measurement results

In our developed STB, the measured distances range from 3 to 30 cm for the fill level, while the threshold for the door open is 10 cm. Hence, we conducted three experiments measuring an object placed 3, 10, and 30 cm away from the sensor. In each experiment, we took 50 consecutive observations, the corresponding dispersion of which is shown in Figs. 7-9. The SD, SDM and points meeting the condition in Eq. (3) are respectively visualized in blue, red, and green. Notably, the SDM curves converge at a faster rate compared to the SD ones, verifying the suggestion in [19] that $s(\bar{q})$ is a useful index to evaluate the precision of the sample mean to the to-be-estimated population mean. Besides, although green points appear early at $n < 10$ in Figs. 7 and 9, the stability is obtained in all three test cases at n starting from 18. We conclude that at $n = 18$, the sample mean is close to the expected population mean and select it as an index to calculate the average distance to feed to the Arduino board.

The obtained results from the above tests were used to verify the function of the STB in the processing of the cover door and the lids according to the motion of users' hands to the STB and the fullness of the bins.

In the door cover experiment, a participant placed their hand in front of the sensor 150 times from a distance of 10 cm in different directions, opposite, left, and right, to trigger the door cover to open. The number of successful cases is 142 out of 150. We observed that the 8 failure cases appeared when the hand was placed out of the $[-22.5, 22.5]$ degree range of the sensor's effective viewing angle cone. The recorded error increased at higher viewing angles, i.e. 5 cm at 30 degrees, leading to the incorrect detection of the system.

Furthermore, we measured the fill level of bins A and B by evaluating the distance from the sensor to the object inside the bins, the range of which is from 5 to 20 cm. Besides, the measurements at different room temperatures, i.e. 21, 25, and 30 degrees, are also reported. Let TV be the conventional true value of the distance to be measured, $\bar{q}^{[t]}$ and $\bar{\epsilon}^{[t]}$ be the mean value of n consecutive records its corresponding mean error to TV . The obtained results are shown in Table I. Notably, the measurements are impacted by the temperature, and the sensor seems to work steadily at 25 Celcius degrees. This information can be used for future development of the STB.

B. Model evaluation

The performance of the selected learning models is evaluated on the Jetson Nano's CPU and GPU. The results of the first test are reported in Table II. It is significant to see that YOLOv10 outperforms its counterparts on all accuracy indices, including F_1 , Acc , Pr , and Rc . On the other hand, MobileNetV2 stands out in inference time, processing almost 3 times faster than the second-best EfficientNetV2 while its F_1 is only lower than that of YOLOv10.

The models are then quantized before being evaluated on the Jetson Nano's GPU, the results of which are reported

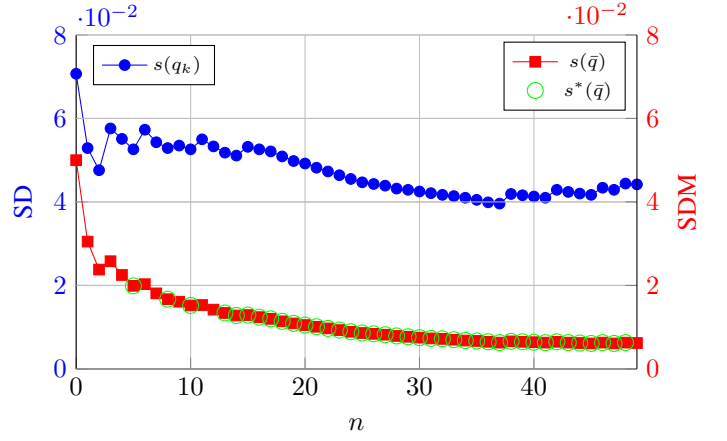


Fig. 7: Corresponding SD and SDM of n at 3 cm.

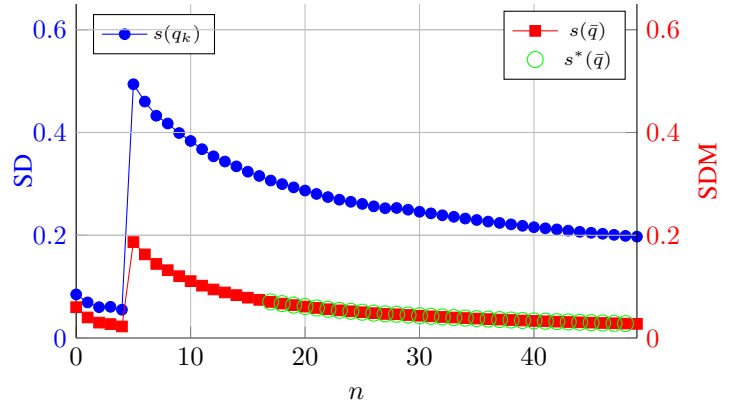


Fig. 8: Corresponding SD and SDM of n at 10 cm.

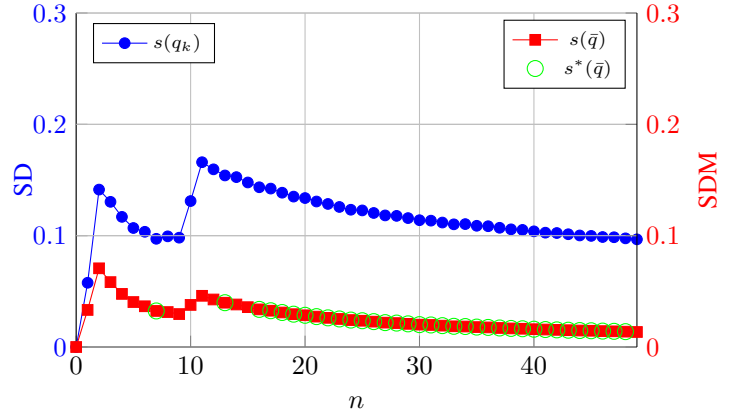


Fig. 9: Corresponding SD and SDM of n at 30 cm.

in Tab. III. Notably, the trade-off between the classification accuracy and inference time is observed on all participating models. In this experiment, MobileNetV2 experienced the best balance between performance and accuracy, i.e. it runs almost 3 times faster while being only 0.55% less accurate in F_1 . Meanwhile, a significant decrease in F_1 is observed on the optimized YOLOv10, from 99.4% to 76.61%, although the

TABLE I: MEASUREMENT ERRORS AT DIFFERENT TEMPERATURES

TV	$\bar{q}^{[21]}$	$\bar{e}^{[21]}$	$\bar{q}^{[25]}$	$\bar{e}^{[25]}$	$\bar{q}^{[30]}$	$\bar{e}^{[30]}$
5	5.26	0.26	4.97	-0.03	4.60	-0.40
10	10.27	0.27	9.96	0.04	9.55	-0.45
15	15.27	0.27	14.89	-0.11	14.66	-0.34
20	20.10	0.10	19.82	-0.18	19.56	-0.45

TABLE II: MODEL PERFORMANCE ON CPU

Model	Acc	F1	Pr	Re	FpS
VGG16	97.72%	97.41%	96.75%	98.18%	0.25
MobileNetV2	99.09%	99.02%	98.79%	99.26%	8.21
EfficientNetV2	97.72%	98.14%	98.14%	98.14%	2.80
YOLOv10	99.54%	99.40%	99.62%	99.18%	1.27

TABLE III: MODEL PERFORMANCE ON GPU

Model	Acc	F ₁	Pr	Rc	FpS
VGG16 (GPU)	79.09%	74.24%	72.31%	84.74%	6.65
MobileNetV2 (GPU)	98.63%	98.65%	98.42%	98.92%	23.65
EfficientNetV2 (GPU)	41.18%	22.63%	35.00%	25.12%	19.72
YOLOv10 (GPU)	78.18%	76.61%	82.22%	81.81%	7.10

inference time has decreased almost 6 times.

IV. DISCUSSION

Although promising results have been observed, there are areas to further cultivate in our future work. The door cover test has a high success rate. However, the viewing angle is currently limited, leading to unwanted results in the processing. In addition, the fill level experiment showed how temperature affects the sensors' measurements. The sensor's performance in future development might be enhanced by resolving these temperature sensitivity problems and angle-related inaccuracies.

The classification results have shown potential in waste management, especially with the applications of lightweight models, such as the MobileNet. The accuracy of YOLOv10 while performing on CPU is also noteworthy. Future developments might include the performance optimization of YOLOv10 or the combination of YOLOv10 and MobileNet to leverage the advantage of each model.

V. CONCLUSION

This work proposed a vision-based waste classification approach, allowing the automatic sorting of domestic waste and reducing human intervention. Experiments on our collected dataset and the developed STB have confirmed the effectiveness of some recent detection models, based on deep learning with neural networks. The learning models, along with the development of a smart trash bin, present a promising approach to the problem of household waste management, addressing environmental sustainability in modern cities. Besides the optimization in performance and accuracy of these models, as well as the mechatronic aspects of the smart bin, our future work will consider increasing the size of our collected dataset in both variety and quantity.

REFERENCES

- [1] Silpa Kaza, Lisa Yao, Perinaz Bhada-Tata, and Frank Van Woerden. *What a waste 2.0: a global snapshot of solid waste management to 2050*. World Bank Publications, 2018.
- [2] Ayesha Siddiqua, John N Hahladakis, and Wadha Ahmed KA Al-Attiya. An overview of the environmental pollution and health effects associated with waste landfilling and open dumping. *Environmental Science and Pollution Research*, 29(39):58514–58536, 2022.
- [3] Peter W Tait, James Brew, Angelina Che, Adam Costanzo, Andrew Danyluk, Meg Davis, Ahmed Khalaf, Kathryn McMahan, Alastair Watson, Kirsten Rowcliff, et al. The health impacts of waste incineration: a systematic review. *Australian and New Zealand journal of public health*, 44(1):40–48, 2020.
- [4] Dietmar Offenhuber. *Waste is information: infrastructure legibility and governance*. MIT Press, 2023.
- [5] Wenqi Wu and Ming Zhang. Exploring the motivations and obstacles of the public's garbage classification participation: evidence from sina weibo. *Journal of Material Cycles and Waste Management*, 25(4):2049–2062, 2023.
- [6] Haruna Abdu and Mohd Halim Mohd Noor. Domestic trash classification with transfer learning using vgg16. In *2022 IEEE 12th International Conference on Control System, Computing and Engineering (ICCSCE)*, pages 137–141. IEEE, 2022.
- [7] Wahied Khawar Balwan, Arashdeep Singh, and Sachdeep Kour. 5r's of zero waste management to save our green planet: A narrative review. *European Journal of Biotechnology and Bioscience*, 10(1):7–11, 2022.
- [8] K Mostaghimi and J Behnamian. Waste minimization towards waste management and cleaner production strategies: a literature review. *Environment, Development and Sustainability*, 25(11):12119–12166, 2023.
- [9] Sana Shahab, Mohd Anjum, and M Sarosh Umar. Deep learning applications in solid waste management: A deep literature review. *Int. J. Adv. Comput. Sci. Appl.*, 13(3), 2022.
- [10] Haonan Fan, Qin Dong, Naixuan Guo, Jun Xue, Rongrong Zhang, Haobo Wang, and Mingfeng Shi. Raspberry pi-based design of intelligent household classified garbage bin. *Internet of Things*, 24:100987, 2023.
- [11] Zhichao Chen, Jie Yang, Lifang Chen, and Haining Jiao. Garbage classification system based on improved shufflenet v2. *Resources, Conservation and Recycling*, 178:106090, 2022.
- [12] Aria Bisma Wahyutama and Mintae Hwang. Yolo-based object detection for separate collection of recyclables and capacity monitoring of trash bins. *Electronics*, 11(9):1323, 2022.
- [13] Yuneza Claire Y Mortos, Mon Arjay Malbog, Marte Nipas, Dennis Nava, et al. Smartsort: YOLOv4-driven smart sorting for household waste management. In *E3S Web of Conferences*, volume 488, page 03025. EDP Sciences, 2024.
- [14] Xiangkui Jiang, Haochang Hu, Yuemei Qin, Yihui Hu, and Rui Ding. A real-time rural domestic garbage detection algorithm with an improved yolov5s network model. *Scientific Reports*, 12(1):16802, 2022.
- [15] Songzhe Pan, Ning Wang, Yuyun Lin, and Jianhao Tang. Based on yolov8 intelligent trash can garbage classification detection algorithm. *Mathematical Modeling and Algorithm Application*, 2(1):28–32, 2024.
- [16] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4510–4520, 2018.
- [17] Mingxing Tan and Quoc Le. Efficientnetv2: Smaller models and faster training. In *International conference on machine learning*, pages 10096–10106. PMLR, 2021.
- [18] Ao Wang, Hui Chen, Lihao Liu, Kai Chen, Zijia Lin, Jungong Han, and Guiguang Ding. Yolov10: Real-time end-to-end object detection. *arXiv preprint arXiv:2405.14458*, 2024.
- [19] Joint Committee for Guides in Metrology. *Evaluation of measurement data—guide to the expression of uncertainty in measurement*, 2008.
- [20] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.
- [21] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *Computer Vision—ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part V 13*, pages 740–755. Springer, 2014.