Reverse Vending Machine Item Verification Module using Classification and Detection Model of CNN

Razali Tomari[1*], Nur Syahirah Razali[2], Nurul Farhana Santosa[3], Aeslina Abdul Kadir[4], Mohd Fahrul Hassan[5]
Institute for Integrated Engineering (IIE) [1]
Faculty of Electrical and Electronic Engineering[2, 3]
Faculty of Civil Engineering and Built Environment[4]
Faculty of Mechanical & Manufacturing Engineering[5]
Universiti Tun Hussein Onn Malaysia
86400 Parit Raja, Batu Pahat, Johor. Malaysia

***Abstract*—Reverse vending machine (RVM) is an interactive platform that can boost recycling activities by rewarding users that return the recycle items to the machine. To accomplish that, the RVM should be outfitted with material identification module to recognize different sort of recyclable materials, so the user can be rewarded accordingly. Since utilizing combination of sensors for such a task is tedious, a vision-based detection framework is proposed to identify three types of recyclable material which are aluminum can, PET bottle and tetra-pak. Initially, a self-collected of 5898 samples were fed into classification and detection framework which were divided into the ratio of 85:15 of training and validation samples. For the classification model, three pre-trained models of AlexNet, VGG16 and Resnet50 were used, while for the detection model YOLOv5 architecture is employed. As for the dataset, it was gathered by capturing the recycle material picture from various point and information expansion of flipping and pivoting the pictures. A progression of thorough hyper parameters tuning were conducted to determine an optimal structure that is able to produce high accuracy. From series of experiments it can be concluded that, the detection model shows promising outcome compare to the classification module for accomplishing the recycle item verification task of the RVM.**

***Keywords*—Convolutional neural network (CNN); classification; detection; reverse vending machine (RVM); You Only Look Once (YOLO)**

I. INTRODUCTION

Malaysia is leaving behind in sustaining waste management awareness, especially in recycling [1]. Currently, the rate of solid waste increase significantly [2] and one of the methods that can be done in managing waste effectively is by boosting recycling activities using interactive Reverse Vending Machine (RVM). RVM works by analyzing every deposit recycle materials to the machine and provide reward to the user accordingly. Previously, a hybrid sensing based RVM [3-4] has been developed and tested in municipal office as shown in Fig. 1. However, it manage to recognize only PET bottle and aluminum can, and required tedious sensor calibration, maintenance and not suitable for long term usage.

To cater such an issue, a vision-based technology is incorporated into the RVM material identification module. Vision system capable to recognizing more types of recycle items with vast amount of the collected data from the sample. One of the work is from [5] in which they introduced ThrashNet dataset and use SIFT feature with SVM and CNN model of AlexNet -like architecture. The former model manage to obtain average of 63% performance while the latter show deteriorate performance with 22% accuracy, in which the author argue more dataset will yield a better result. Andrey et al. [6] developed a reverse vending machine with several CNN classification model by analyzing effect of training by combining two different dataset cluster during training. In average the CNN model can produce more than 85% accuracy. They later on test the module in real implementation by combining weigh sensor with the CNN for fraud detection [7].

Recently, CNN becomes trends for thrash items classification either using standalone model, combine with conventional classifier or using ensemble CNN architecture. RecycleNet [8] was introduced to exhaustively analyze optimal state of the art CNN structure for the RVM classification task. They exhaustively tune the model based on empty-structure model, with pre-trained and fine tuning. In average for the performance wise, 90% accuracy can be obtained for each models, layer modification is necessary to ensure processing time and accuracy can be well balance. In [9], a combination of GoogleNet with SVM show promising outcome with 97.86% based on TrashNet dataset. Apart from that a MobileNet variants [10] shows at par performance with 96.57% and optimized DenseNet121 model obtain 99.6% accuracy [11] using the same dataset. An ensemble based model that combine GoogleNet, ResNet-50 and MobileNetv2 using unequal precision measurement data fusion been tested using ThrashNet and FourThrash dataset [12], apparently they claim that the combination provide more robust results during data aggregation of the CNN forecasting results.

Fig. 1. Example of RVM using Combination of Sensors.

There are many works that previously focus on classification models and detection models for the RVM application. However, to the best of our knowledge no comparison done to investigate the effectiveness between each of the models. In this paper a CNN classification based model is compare with detection based model to determine optimal structure for RVM implementation. The paper is organize as follow: section II will discuss about methodology use through this project, follow by result and discussion in section III and eventually project conclusion in section IV.

II. Models and Method

In this section, detail explanation about model and architecture used throughout this project is thoroughly explained. It comprises of three main subsections, namely, dataset preparation, classification model development and detection model development.

A. Datasets Preparation

The arrangement of getting sample image of all classes are as in Fig. 2 in which the camera is vertically locate 21cm above the ground and the sample was placed 51cm from the camera. There are total of 5898 samples collected ranging from three categories of aluminum can, PET bottles and tetra-pak as depicted in Fig. 3. The collected images are then separated into training and validation cluster with a portion of 4961 samples for validation and 937 samples for validation. Details of sample distributions for each cluster can be seen in Table I. It can be noticed that. For the detection task, all images must undergo an annotating process in which in this project a LabelImg software is used.

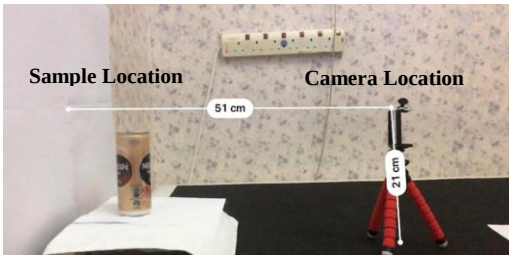

Fig. 2. Setup Arrangement for Dataset Preparation.

Fig. 3. Sample of Tetra-Pak, PET Bottle and Aluminium Can.

TABLE I. Number of Image Dataset in each Recycle Item Image Category

Images	Training	Validation
Aluminum Can	1335	248
PET Bottle	3018	569
Tetra-pak	608	120
Total	4961	937

B. Classification Model Development

For the classification module, a Convolutional Neural Network (CNN) based model is employ. Basically, CNN comprises two main part which are feature extractor module, also known as convolutional layer, and classification module that will regard as dense layer. The former ensure local features from inputted image can be highlighted, while the latter utilize the extracted local features to identify the trained object inside the image. In this project, a pre-trained CNN model is utilize to recognize the three cluster of the recyclable items. It means that, the parameters from convolutional layer of state of the art CNN model will be reuse to be trained with our own dataset. During the training session, convolutional layer parameters will be frozen while dense layer parameters will by dynamically adjusted to reduce the cost function error.

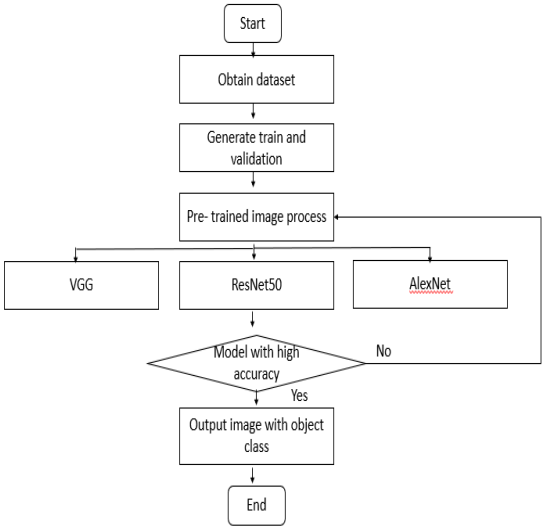

Fig. 4. Block Diagram flow of the Classification Procedure.

Fig. 4 depicted overall system flow of the classification part. From the collected sample set, the training cluster data will be tested with three pre-train models which are VGG, ResNet50 and AlexNet. All the three models will undergo rigorous hyper parameters tuning to analyses the best model that can optimally classify the three types of the recyclable items. In this paper, a free GPU from Google Colab is employed to execute the training process. This platform is a cloud based of Jupyter notebook and can be edited by the team members and easy to access without requiring any setup. It supported many types of machine learning libraries and the sample data can be simply loaded from the user Google drive.

The CNN VGG16 model [13] is the first model that will be access for the classification module. It comprises of systematic architecture of 3x3 filter throughout the 16 layers and was introduced in 2014 as an improvement of Alexnet. The model achieved 92.7% accuracy performance in ImageNe dataset and consists of 138 million parameters and make it a bit slow to train. Fig. 5 shows an architecture of VGG16 model. All the input images will be resized to 224x224 dimension prior to feed to the 13 convolutional layers. Number of channel in this layer is started with 64 channels and was incremented to the factor of two up until 512 channels. In between of the convolutional layer there are five pooling layers that

responsible to down sample an image and was done using maxpooling operation of 2x2 kernel with stride of two. The last convolutional layer outputting feature map with the size of 7x7x512. Then, this 2D features is flatten through the three fully connected layers with ReLu activation in the first two hidden layer and softmax in the output layer.

Residual Network (ResNet) [14] is the second model that will be tested for this project. ResNet is the first model that can be used to generate very deep network from hundreds to thousands layer while still providing a good performance. It able to handle such situation by introducing residual block with a skip connection, which will add a result from previous layer to the next layer of the model as shown in Fig. 6. A very deep network literally will suffer from vanishing gradient problem, in which the back propagate training error value will become smaller from layer to layer and eventually becomes zero. In the layer where the error goes to zero and in its subsequent, no more parameters update will be executed and hence the model unable to converge well with the given data. ResNet handle such situation via the skip connection which allowing the gradient to flow via the alternative path. By doing that, it will ensure the higher layer will perform as good as the lower layer.

In this project, a ResNet50 model is used in which it contains 49 convolutional layers and single dense layer as shown in Fig. 7. The former layer can be further categorize into four main blocks of conv2, conv3, conv4 and conv5 in which each block contain three convolutional layers as shown in Fig. 6. The number of channel in each block is sequentially increment by a factor of two and the model is expected an input image of 224x224 dimension. In summary, the convolutional structure can be visualize as: conv1; 3 x conv2; 4 x conv3; 6 x conv4; 3 x conv5. After series of convolutional layers, a dense layer with 1000 neurons and softmax activation function is add up to classify the dataset cluster accordingly.

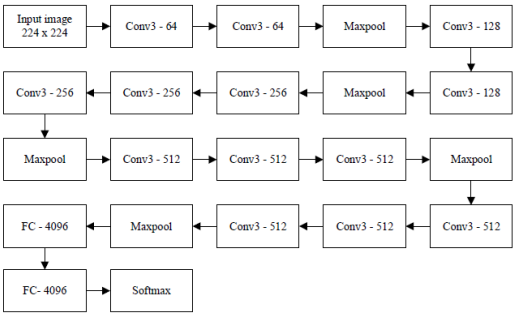

Fig. 5. VGG-16 Structure.

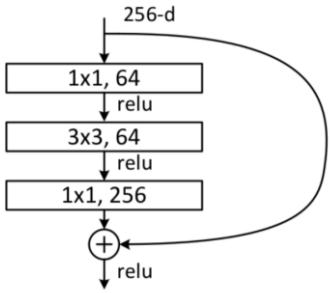

Fig. 6. Residual Block Structure.

layer name	output size		50-layer
conv1	112×112	7×7, 64, stride 2	
conv2_x	56×56	3×3 max pool, stride 2	[1×1, 64; 3×3, 64; 1×1, 256] ×3
conv3_x	28×28		[1×1, 128; 3×3, 128; 1×1, 512] ×4
conv4_x	14×14		[1×1, 256; 3×3, 256; 1×1, 1024] ×6
conv5_x	7×7		[1×1, 512; 3×3, 512; 1×1, 2048] ×3
	1×1	average pool, 1000-d fc, softmax	

Fig. 7. ResNet-50 Structure.

The third model that will be analyze is AlexNet architecture [15]. This is one of the simplest and earliest CNN architecture that have firm grip of overfitting problem by introducing data augmentation and dropout layer. Basically AlexNet consist of eight layers that compose of five convolutional layers and three dense layer as shown in Fig. 8. The model also introduce rectify linear unit (ReLU) and max pooling layer to be use along with the convolutional layer. Regarding the input, it must be resize to 227×227 dimension prior feed to the network. In their structure, three size of filter dimensions were utilized which in dimension of 11×11, 5x5 and 3x3.

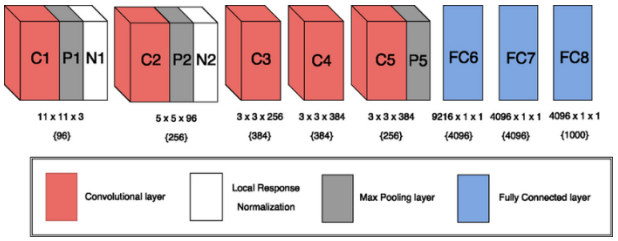

Fig. 8. AlexNet Structure.

To utilize all the mentioned state of the art CNN models in this project, a transfer learning process is implemented. Technically, the process will re-utilize all the convolutional parameters from the model, omit its original dense layer and then hook up our dense layer as shown in Fig. 9. In the figure, shade block area denotes the convolutional layer parameters and will be frozen during training session. On the other hand, the white block area is the new dense layer that consist of two hidden layers with 512 and 128 neuron and one output layer with three neurons that denote our three clusters of aluminium can, PET bottle and tetra-pak. The dense layer, will be dynamically updated during the training session based on the back propagate error value from the output layer.

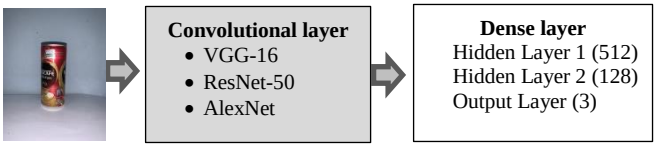

Fig. 9. Visualization of Structure of Transfer Learning Process.

C. Detection Model Development

For detection module, the main structure employ in this project is based on 'You Only Look Once' (YOLO) object detection. Basically, it starts with the forming of grid cells, follow by class prediction across scales, and eventually bounding box location estimation via regression process. The finer grid cell enable for smaller target detection and anchor box make it possible to detect an overlapping object with high accuracy. There are many variants of YOLO model starting from version v1 to version v5 [16-20]. In this project, the recent model which is YOLOv5 is used.

YOLOv5 is a single stage object detector and has three components which are backbone, neck (PANet) and head (output) as depicted in Fig. 10. Model backbone is mostly used to separate significant features from the input image using cross stage partial network (CPSNet) [21]. CSPNet has demonstrated huge improvement in processing time with more profound networks and solves the problems of repeated gradient information in large-scale backbones. Such structure will improve inference, speed, and accuracy and at the same time reduce the model size. In RVM verification task, detection speed, and accuracy is important, and compact model size also determines its inference efficiency on compact device controller.

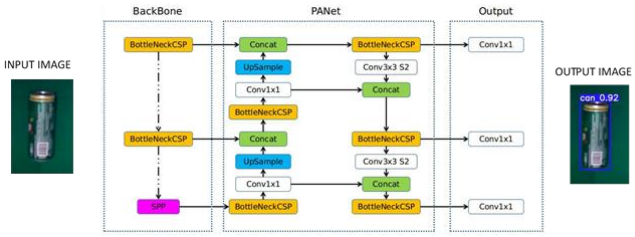

Fig. 10. YOLOv5 Architecture.

Model neck is mostly used to produce feature pyramid to assist models to deduct on object scaling. It assists with recognizing similar item with various sizes and scales. Feature pyramids are extremely valuable and assist models with performing admirably on hidden data. In YOLOv5 path aggregation network (PANet) [22] is utilized as neck to get feature pyramids. PANet improve localization signals accuracy the in lower layers, and hence enhance the location accuracy of the object.

The model head is essentially used to play out the last detection part. It applied three different size of anchor boxes to cater small, medium and big objects on features, and creates final output vectors with class probabilities, object scores, and bounding boxes. The YOLOv5 model head structure is equivalent to the past YOLOv3 and YOLOv4 model.

YOLOv5 has four final architecture which are YOLOv5s (small), YOLOv5m (medium), YOLOv5l (large) and YOLOv5x (xlarge). To balance between speed and accuracy of the system, YOLOv5s which is the smallest and fastest model is utilised for the RVM implementation.

In YOLO family the cost function calculation is source from objectness score, class probability score, and bounding box regression score. YOLOv5 has utilized binary cross-entropy (*BCE*) with logits loss equation from for loss calculation of class probability and object score. This loss joins a sigmoid layer and the *BCE* loss in one single class and more mathematically stable than utilizing a plain sigmoid followed by a BCE loss.

III. Results and Discussion

In this section performance of the selected classification and detection model is presented. Basically the section comprises of three main parts which are classification model assessment, detection model assessment and optimal model assessment in live feed video streaming. Part of the samples that will be use during the training and validation session can be seen in Fig. 11 which shown PET bottle samples, aluminum can samples and tetra-pak samples.

A. Classification Model Assessment

For the first classification analysis, summary of the outcomes can be seen in Fig. 12. It displays result of training session, validation session and loss in every epochs. Initially, in every training session 50 epoch will be selected, and the location where overfitting start to occurs will be selected as the new epoch for the next training session. The tuning process will be repeated until optimal outcome that well balance between training and validation data is obtained. However, if under fitting condition constantly occurs, then it can be concluded that the model cannot be used for recognizing the given dataset.

As can be seen in figure, the VGG-16 model performance keep under fitting both for training and validation with value of below 70%, and hence it can be summarize that the VGG-16 model unable to works well to classify the data. Next, for the AlexNet model it begins to under fit after 15 epochs and eventually at 50 epochs the performance for training is at 98% while for the validation is at 80%. It can be say that, the ALexNet model works well for memorizing the data but not works well for recognizing unseen data pattern. Finally, for the ResNet-50 model, it can be observed that it able to obtain high accuracy during training session with 92%, but the validation performance fluctuate significantly with the lowest value of 12% after 50 epochs.

Fig. 11. Sample of Training Images.

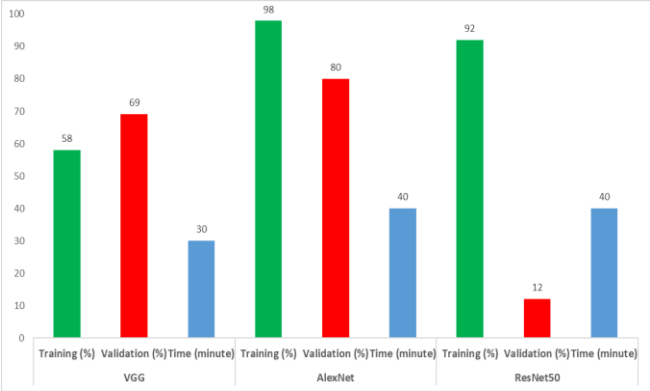

Fig. 12. Summary of Classification Model Analysis.

Based on the finding, for RVM classification model the best structure that can balance trade-off between training and validation is the AlexNet model with 80% performance during validation stage and the highest training performance among other two models. Sample of the classification outcome can be seen in Fig. 13.

Fig. 13. Sample of Classification Module Outcome.

B. Detection Model Assessment

This section will deliberately discuss result obtained for the YOLOv5 detection module framework. The best variable must be selected to achieve a rapid object detection model with high accuracy. There are three important parameters that are frequently used for object detection assignment: Generalized Intersection over Union (GIoU) graph, Objectness graph and mean Average Precision (mAP) graph.

GIoU graph indicates how near the ground truth with the predicted bounding box, whereas for Objectness graph means confidence score whether there is an object in the grid cell. In the event that the bounding box covers a ground truth object more than others, the relating Objectness score ought to be 1. The third graph that is important is mAP graph where the precision and recall for all the objects introduced in the images ought to be figured. It additionally needs to consider the confidence score for each object detection by the model in the picture. Consider the entirety of the predicted bounding boxes with a confidence score over a specific limit. Bounding boxes over the set threshold value are considered as positive boxes while all predicted bounding boxes underneath the set threshold value are considered as negative. Thus, the higher the threshold value is, the lower the mAP will be, however the confidence is more accurate.

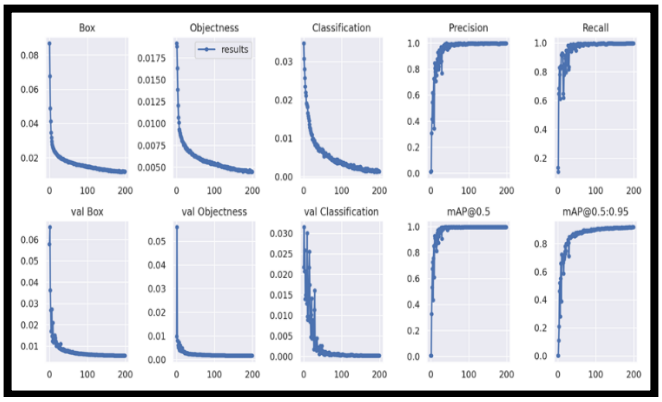

Fig. 14. Detection Training Result of Fine Tuning Process for 200 Epochs.

Fig. 14, summarize the tuning outcome after 200 epochs of training. The first three columns show loss function value for GIoU confident score, Objectness confidence score and classification confidence score, where the upper part denoted the training data while the lower part the validation data. It can be observes that all the loss value decrement constantly near to zero for the three scores. Even there is some fluctuation for the val-classification score in the beginning of training, it becomes constantly goes to zero as the epoch more than 50. Looking at another four remaining curve figures of precision, recall, mAP@0.5 and mAP@0.5: 0.95, the performance gain high confidence score after 100 epochs.

To gain more detail insight of each object class performance, precision and recall curve can be investigated and is shown in Fig. 15. Precision recall curve shows the tradeoff between the exactness and recall esteems for various limits. This curve assists with choosing the best threshold to expand both measurements. As the number of positive samples get higher (high recall), the accuracy of classification of every sample precisely get lower (low precision). From the figure, it can be observed that, the PET bottle shows highest confidence score with 0.997, follow by tetra-pak and finally the aluminium can. In average, overall performance of all class detection is at 0.995 using 0.5 confidence score of mean average precision. The performance for the detection can be conclude as much more better than the one obtain from the classification model based on the given dataset. The resultant images after running the inference/testing process were as in Fig. 16.

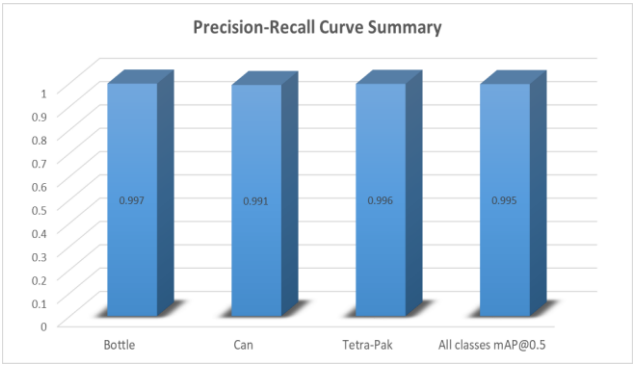

Fig. 15. Precision Recall Curve of the YOLOv5 Model.

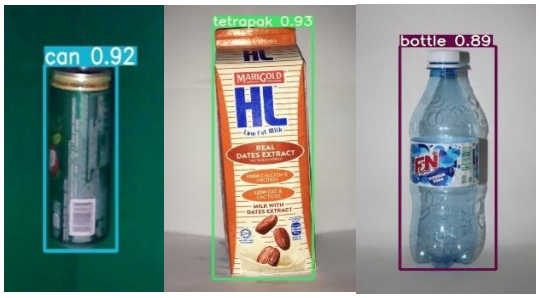

Fig. 16. Sample of YOLOv5 Detection Outcome.

C. Optimal Model Assessment with Live Feed Video

The live feed video application is done to gain insight how well the model works under real implementation scenario of the reverse vending machine. Basically the procedure will include the utilization of the webcam and the recyclable items will be slide within the camera field of view. Since based on the two model analysis the detection based architecture show a better outcome, analysis in this section was done for the detection module part only.

For the first assessment 10 samples for each class which total up to be 30 were feed in the webcam view while the detection algorithm running. The finding is PET bottle class has achieved 100% accuracy as the module correctly detects all of the input that has been fed simultaneously, whereas for aluminium can class, the module only achieved 80% accuracy and for the tetra-pak class 90% accuracy. Based from this outcome, it can be concluded that the module gained average of 90% accuracy in real implementation condition. Sample of the snapshot during experiment is shown in Fig. 17.

Fig. 17. Sample Snapshot during Live Feed Testing of the Detection Module.

Fig. 18. Result after the Detection Module has been Tested under Condition of more than Two Classes in one Frame.

For the second condition, samples of recyclable waste material will be put together in front of the webcam. The aim is to analyze module capability for detecting multiple samples in single feed of camera image. The result will be evaluated whether the module can detect the cluster precisely. Result obtained in this condition is recorded in Fig. 18 and it can be concluded that the module works well to fulfill the task by successfully detecting all the recycle materials given.

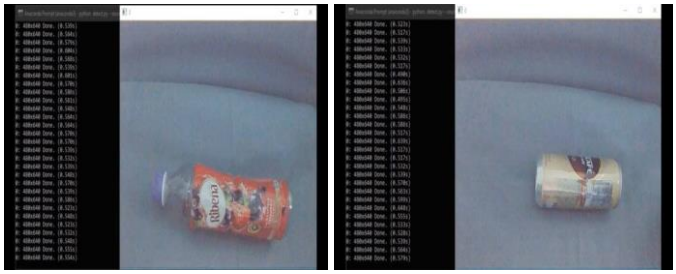

Fig. 19. Detection Result with Low Light Condition.

Finally, to imitate the real condition situation inside RVM, the detection module is tested under low light condition with only exposure from the natural light. The result can be seen in Fig. 19. This shows that, without proper lighting the module unable to works well and hence light source is crucial to guarantee system success. As for the computational cost during live feed assessment, based on the analysis of feeding the detection module with 314 images, it is found that in average the processing time is 482.1 millisecond for color image with dimension of 640x640. The processing speed is acceptable for RVM usage and can be further improve by incorporating GPU based controller such as NVIDIA Jetson Nano or by using OpenCV OAK-1 camera.

IV. Conclusion

In this project, a CNN-based classification and detection module were investigated to be used as RVM verification module. Three class of sample which are PET bottle, aluminium can and tetra-pak are used during the training and validation stage with amount of 4961 and 937 respectively.

The highest performance model either from classification or detection was then undergo live feed video assessment under RVM implementation condition.

For the classification models, three CNN architecture were tested which are VGG-16, ResNet-50 and AlexNet. From series of training and fine tuning, it can be concluded that AlexNet model show high performance with 98% accuracy during training and 80% during validation stage, follow by ResNet-50 and then VGG-16. As for the detection model, YOLOv5 is used and it shows promising outcome with average of 99.5% mAP@0.5 accuracy based on the give training and validation data. Since the detection model show promising outcome compare to the classification model, it was further tested under RVM real implementation condition using a life feed webcam data. From the testing it can be concluded that, the system able to gain 90% accuracy during testing and apparently a good source of light is crucial to ensure the successfulness of the detection process.

The finding in this project is subject to several limitations that could be addressed in future research. First, the system is currently tested in a lab condition that simulating RVM functionality and hence future assessment in actual operation condition is in planning for system assessment. Second, most of the collected sample items were gathered locally and hence there is tendency that the system unable to detect recyclable items from beyond local brand. Finally, the detection model is currently tested with state of the art YOLO detector, in future, other type of detection algorithm such as single shot detector and RCNN variants can be assess to compare its outcome with the YOLO based platform.

ACKNOWLEDGMENT

This research is supported by Universiti Tun Hussein Onn Malaysia (UTHM) through Multidisplinary Research Grant Scheme (MDR) (Vot. No. H495).

REFERENCES

[1] Jereme, I. A., Siwar, C., & Alam, M. M. (2015, October). "Waste recycling in Malaysia: Transition from developing to developed country", Indian Journal of Education and Information Management, 4 (1), pp. 1- 14, 2015.

[2] N.A. Mokhtar, "Malaysia masih ketinggalan dalam amalan kitar semula",Berita Harian Online, (2016, October 25).

[3] R. Tomari, A. A. Kadir, W.N.W Zakaria, M. F. Zakaria. M.H.A Wahab & M.H. Jabbar, "Development of Reverse Vending Machine (RVM) Framework for Implementation to a Standard Recycle Bin", Procedia Computer Science, vol. 105, pp. 75-80, 2017.

[4] R. Tomari, M. F. Zakaria , A. A. Kadir, W.N.W Zakaria, M.H.A Wahab ,"Empirical Framework of Reverse Vending Machine (RVM) with Material Identification Capability to Improve Recycling", Applied Mechanics and Materials, pp. 114-119, 2019.

[5] M. Yang and G. Thung, "Classification of trash for recyclability status", CS229 Project Report 2016, 2016.

[6] A. N. Kokoulin, A. I. Tur and A. A. Yuzhakov, "Convolutional neural networks application in plastic waste recognition and sorting," 2018 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus), Moscow, 2018, pp. 1094-1098.

[7] A. N. Kokoulin and D. A. Kiryanov, "The Optical Subsystem for the Empty Containers Recognition and Sorting in a Reverse Vending Machine," 2019 4th International Conference on Smart and Sustainable Technologies (SpliTech), Split, Croatia, 2019, pp. 1-6,

[8] C. Bircanoğlu, M. Atay, F. Beşer, Ö. Genç, & M. A. Kızrak, "RecycleNet: Intelligent Waste Sorting using Deep Neural Networks", In 2018 Innovations in Intelligent Systems and Applications (INISTA). 2018, pp. 1-7.

[9] U. Ozkaya and L. Seyfi, "Fine-Tuning Models Comparisons on Garbage Classification for Recyclability", arXiv preprint, arXiv:1908.04393, 2019.

[10] Z. Dimitris, T. Dimitris, B. Nikolaos and D. Minas, "A Distributed Architecture for Smart Recycling Using Machine Learning", Future Internet, 12, 141, 2020, pp 1-13.

[11] W.-L. Mao, W.-C. Chen, C.-T. Wang, Y.-H. Lin, "Recycling waste classification using optimized convolutional neural network", Resources, Conservation and Recycling, vol.164, 2021, 105132,

[12] H. Zheng and Y. Gu," EnCNN-UPMWS:Waste Classification by a CNN Ensemble Using the UPM Weighting Strategy", Electronics, 10, 427, 2021.

[13] K. Simonyan, A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition", arXiv preprint, arXiv: 1409.1556, 2014.

[14] K. He, X. Zhang, S. Ren, J. Sun, "Deep Residual Learning for Image Recognition", arXiv preprint, arXiv: 1512.03385v1, 2015.

[15] A. Krizhevsky, I. Sutskever, G.E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks", In Proc. of 25th International Conference on Neural Information Processing System, vol. 1, 2012, pp. 1097-1105.

[16] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 2016-Decem, pp. 779–788, 2016.

[17] J. Redmon and A. Farhadi, “YOLO9000: Better, faster, stronger,” Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017, vol. 2017-January, pp. 6517–6525, 2017.

[18] J. Redmon and A. Farhadi, “YOLOv3: An Incremental Improvement,” arXiv preprint arXiv:1804.02767, 2018.2018.

[19] A. Bochkovskiy, C.-Y. Wang, and H.Y. M. Liao. Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934, 2020.

[20] G. Jocher, A. Stoken, J. Borovec, NanoCode012, A. Chaurasia, TaoXie, L. Changyu, Abhiram V, Laughing, tkianai, yxNONG, A. Hogan, lorenzomammana, AlexWang1900, J. Hajek, L. Diaconu, Marc, Y. Kwon, oleg, wanghaoyang0106, Y. Defretin, A. Lohia, ml5ah, Ben Milanko, Benjamin Fineran, Daniel Khromov, DingYiwei, Doug, Durgesh, andFrancisco Ingham. ultralytics/yolov5: v5.0 -YOLOv5-P6 1280 models, AWS,Supervise.ly and YouTube integrations, Apr. 2021

[21] C.Y. Wang, H.Y. Mark Liao, Y.H. Wu, P.Y. Chen, J.W. Hsieh, I.H. Yeh, “CSPNet: A new Backbone that can Enhance Learning Capability of CNN”. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2020), June 2020; pp. 390–391.

[22] K. Wang, J.H. Liew, Y. Zou, D. Zhou, J. Feng, “ PANET: Few-shot image semantic segmentation with prototype alignment”, In Proceedings of the IEEE International Conference on Computer Vision (ICCV 2019), Seoul, Korea, 2019,pp. 9197– 9206.