Real-Time Plastic Surface Defect Detection Using Deep Learning

Muhammad Izzat Bin Roslan
Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450, Shah Alam, Selangor, Malaysia,
E-mail: muhdizzatroslan94@gmail.com

Zaidah Ibrahim
Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450, Shah Alam, Selangor, Malaysia,
E-mail:zaidah@tmsk.uitm.edu.my

Zalilah Abd Aziz
Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450, Shah Alam, Selangor, Malaysia,
E-mail: zalilah@tmsk.uitm.edu.my

2022 IEEE 12th Symposium on Computer Applications & Industrial Electronics (ISCAIE) | 978-1-6654-8703-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/ISCAIE54458.2022.9794475

***Abstract*— Quality control is a process utilized in the plastic packaging industry to ensure that the products that are produced are high-quality. This is achieved by identifying and eliminating defects before they are commercialized in the market. The quality of plastic surfaces makes a difference in how customers see the final product. To avoid experiencing errors and minimize product defects, manual surface defect detection is typically performed by humans through the naked eyes. Due to slow detection speed, high labor costs, and visual acuity limitations, manual defect detection can no longer meet today's demands. Therefore, real-time identification of plastic surface defects using computer vision technology is required. This paper proposes a method for the real-time detection and classification of plastic surface defects using deep learning which is You Only Look Once (YOLO). YOLO has shown excellent performance in object detection and this research applies YOLOv5. It is performed by training a custom dataset obtained from the plastic packaging industries to identify defective surfaces and at the same time to obtain its detection accuracy in terms of precision, recall, F-measure, and mAP.**

Keywords— object detection, plastic surface defect detection, YOLOv5, Deep Learning.

I. INTRODUCTION

Plastic packaging products have been massively produced throughout the century. Almost 380 million tons of plastic products are produced every year. Global plastic production recorded 407 million tons of plastic packaging has been produced throughout the year 2015 which is nearly 200-fold of production compared to the 1950s [1]. This shows how demanding plastic usage can be. Although plastic packaging is considered a major cause of toxic pollutants in the world, it also brings great benefits to the distributor, retailer, and consumer [1]. Industries such as food & beverage, health care, cosmetic and personal care, and consumer goods mostly rely on plastic packaging to contain their product mainly to help prevent contamination, prolong shelf life, prevent waste, display product information, and provide efficient transportation.

As more plastic packaging is being produced due to high demand, manufacturers need to take vigorous steps on complying with the quality standards to the customer and consumer. Factories with a high production rate can average around two or more million bottles, caps, and container products in a day, and within the amount of quantity that has been manufactured at least 1% or more of the product are defective [2]. Defective products may be caused by several internal and external issues during production. Machines with limitations are contributing factors to this issue. With that in mind, formal segregation of defective products needs to be done by identifying the features and characteristics of the defects. In plastic packaging, defective parameters are often identified on the surface of the plastic. Defects such as scratches, black spectacles, die lines, watermarks, and pinholes are some of the defects that might appear on the surface of the plastic during production. These defects are gathered and segregated by the operator during the online inspection by checking the quality one by one repetitively throughout the process. However, the defective features on the surface of the plastic may vary differently from each other. This requires an extensive amount of time to train the operator to be able to detect the defective product especially when the defective features are difficult to be identified such as scratches and black spectacles. Furthermore, enforcing manual labor to check the quality is deemed to be inconsistent resulting in slow detection, human error, and bad judgment.

Due to highly repetitive and labor-extensive tasks in a demanding industry, the traditional method of performing a task is less effective [3]. Therefore, a different method of conducting quality control needs to be introduced by implementing the deep learning technology. Deep learning falls under the Artificial Neural Network (ANN) approach that gives the ability to identify and differentiate objects or defects by training the model. Camera Vision is typically used in defect detection especially in identifying defective features and obtaining the data sets for training. Small defects such as scratches and black spectacles work well-using camera vision due to the nature of the defect, which is relatively small and difficult to detect, hence often overlooked by the human eye. Throughout the years deep learning technology has grown from the Convolutional Neural Network model (CNN) to Recurrent Neural Network (RNN), Long Short-Term Memory Network (LSTM), Generative Adversarial Network (GAN), and several more as each of them has its different functionality in detecting an object.

In this paper, real-time plastic surface defect detection is developed using the deep learning method. It uses the current YOLO v5 model as a classifier to detect and recognize the surface defect on a plastic surface. A dedicated camera vision is used to capture defective features to create a custom dataset which later is used to train the model. In addition, several image augmentation techniques are also used to create a robust dataset to improve detection accuracy. Furthermore, the performance of the model is evaluated in terms of its precision, recall, F-measure, and mAP. In conclusion, as more people are coming into the industry, business owners must compete with other businesses by providing better quality products, services, and customer demand on time. With the approach of utilizing deep learning and computer vision systems in factories, it will help improve the quality checking process greatly, thus being able to meet the quality standard from customers and consumers.

II. LITERATURE REVIEW

A. Convolutional Neural Network (CNN)

The concept of a Convolution Neural Network (CNN) is a multi-layer Artificial Neural Network (ANN) with two-dimensional input data [4]. Each layer of the network is made up of many two-dimensional planes, each of which contains multiple neurons. In this composition, neighboring two layers of neurons are connected, but neurons in the same layer are not coupled. The capacity of the model may be modified by modifying the depth and width of the network, and it has a strong assumption for real images (statistical smoothness and local correlation). CNN uses a weight-sharing network structure to make it more akin to a biological neural network. As a result, CNN can effectively lower the network model's learning complexity, have fewer network connections, and weight parameters, and are more likely to be trained than a fully connected network of significant size.

B. Faster Region-Based Convolution Neural Network (Faster R-CNN)

To tackle the cost issue in the traditional technique that uses selective search methods for the development of region proposals, S. Ren et al. created a novel method known as Region Proposal Network (RPN) to recognize objects, forecast bounding boxes, and generate region proposals in 2016 [4]. As a result, Faster Region-Based CNN is computed using a mix of RPN and Fast Region Based CNN models. In this case, the complete image is used as an input to produce the feature maps. Sliding a 3x3 size window over an entire feature map generates a feature vector that is tied to two fully connected layers. There are two completely linked layers, one for box regression and the other for box classification. Many regional proposals are found using fully connected layers. When the greatest limit of k regions is made permanent, the size of the box-regression layer output is 4k and the size of the box-classification layer output is 2k. The term "anchors" refers to a k region proposal that is discovered via a sliding window. The score was chosen after detecting the anchor boxes, which are only those helpful boxes that are formed by implementing a threshold over the objectless. The core CNN model generates feature maps and anchor boxes, which are used as inputs for the Fast R-CNN model [5].

C. You Only Look Once (YOLO)

For detecting bounding boxes and forecasting their probabilities, the network model is used. It is straightforward to use because it allows for real-time forecasts. The YOLO technique works by taking a full image as an input and fragmenting it into an SxS network, where individual cells of this network are used to forecast B bounding boxes and their confidence scores. The confidence score is calculated by multiplying the probability of identifying objects by the value of IoU within the ground truth boxes.

The Google Net CNN model is used in YOLO, and the inception module is proposed [6]. It consists of two fully linked layers preceded by 24 convolutional layers, with the inception module replaced with a 3x3 convolutional layer preceded by a dimensionality reduction layer, as well as any of the filters: 1x1, 2x2, or 3x3. There are five versions of YOLO (v1, v2, v3, v4, and v5), with the most recent YOLO v5 being the quickest, containing nine convolutional layers and several filters. The output generated by the last layer in YOLO for predicting every cell of the grid is S*S*(C+B*5), where S*S is the total size of the grid, C represents the number of estimated probabilities for each class, and B represents the number of anchor boxes per cell (these boxes are connected to confidence score and 4 coordinates). The ImageNet dataset is used for classification, and it is pre-trained with more than half of the convolutional layers. When using older methodologies, most of the time the objects are discovered by using bounding box prediction. There are many bounding box predictions in the YOLO technique that do not contain the object. The Non-Maximum Suppression (NMS) technique is used at the network's end to tackle this problem. It combines the extremely similar bounding boxes into a single box for identical items, however, there is still a case for false-positive detection.

III. RELATED WORK

Machine learning and deep learning have become more prevalent in the field of computer vision nowadays. This technological advancement has allowed many researchers to develop new algorithms for detecting surface defects. As surface detection has become more and more complicated, researchers are working their best to find out a suitable way to improve its detection accuracy. Hence, it is important to find out which algorithm suits best in detecting surface defects due to the unpredictable number of outcomes.

A. Fabric Surface

The concern in the textile production industry is to ensure the textiles that have been produced in the factory achieve high quality [7]. Due to the shortcomings in manual defect detection methods, the researchers have proposed a Cascaded Faster R-CNN network to classify fabric defects. The Cascaded Faster R-CNN operates by using an Inception-ResNet-v2 network as a pre-classifier which is then sent to the Faster R-CNN network for defect detection. In addition, the author also includes the optimization of NMS (Non-Maximum Suppression) to minimize the redundancy of overlapping detection frames of defected areas.

Another method of detecting fabric defects proposed by the researcher An et al. [7] is the improved Faster R-CNN. The model includes a deep residual network for feature extraction, a multi-scale fusion to achieve small object detection, and SoftMax regularization to improve network convergence ability and classification accuracy. As a result, the researchers have achieved great success in implementing small object detection for fabric defects. Other than that, Liu et al. [8] has proposed a real-time fabric defect detection using Lightweight CNN. The issue with deep CNN is that it heavily relies on computational and storage services. However, by implementing the lightweight CNN the researchers were able to reduce the computational complexity and reduce storage usage while being able to obtain high detection accuracy. Furthermore, Li et al. [9] has proposed fabric detection using the improved Cascaded R-CNN. It uses the Feature Pyramid Networks (FPN) which contain multiple convolution layers to extract defective features within each layer which is then able to obtain a higher detection rate.

B. *Metal Surface*

Metal is a crucial basic material for associated planar sectors such as architecture, aerospace, machinery, and automobiles. Some metal or steel production is known to be complicated, with high standards, and is prone to defects owing to mechanical, human, or environmental factors [10]. To meet the quality of metal surfaces the implementation of deep learning technology plays a huge role in quality assurance. Imposed on that, Xian et. al [11] have proposed an automatic metallic surface defect detection and recognition with CNN. It uses a new Cascaded Autoencoder (CASAE) for segmenting and localizing faults. Whereas the cascade network uses semantic segmentation to translate the input defect image into a pixel-by-pixel prediction mask and a compact CNN is used to classify the defect regions of segmented data into their respective group. Using the industrial dataset, the performance score of the proposed method has been obtained at 89.60 percent in detection accuracy on the metal surface. However, even though the result obtained is good, there is still room for improvement that can be made to increase the detection accuracy for real-time application. Ren et. al [12] proposed a slighter Faster R-CNN for real-time detection of steel strip surface defects where the convolutional layers for feature extraction were replaced with depth-wise separable convolutions, resulting in a three to four-fold boost in network speed. Then, to increase the network's capacity to discern between different sorts of errors, the center loss was added to the original loss function. As a result, the proposed networks have achieved 98.32 percent accuracy with an average speed of 0.05s per image. This shows a great improvement in detection accuracy by using Faster R-CNN compared to Tao et. al [11].

In addition, similarly to the work conducted by both researchers, instead of using the CNN and R-CNN, Li et. al. [13] have proposed a real-time defect detection of steel strip surface by using an improved YOLO detection network method. The YOLO network was upgraded and made entirely convolutional. The proposed upgraded network has 27 convolution layers, and it provides a complete solution for detecting surface flaws in steel strips. With the network, the author examined the six types of problems and achieved mean average precision (mAP) of 97.55 percent and a recall rate of 95.86 percent. Furthermore, with a speed of 83 Frames Per Second (FPS), the network has achieved a 99 percent detection rate.

C. *Irregular Surface*

Irregular surface features such as ceramic surface, fine-ground, metal surface, and solar panels normally have a reflective property which is more difficult to see when the photograph is taken. Due to its reflective and bright physical propert, researchers are trying to understand how to improve the defect accuracy on the surface defect. Based on the work conducted by Zhou et al. [14], they proposed the detection of Micro-Defects on an irregular surface by using the improved Faster R-CNN method. The author has selected the K-means algorithm technique to identify the aspect ratio of the anchor box and uses the feature matrices to the fused different receptive fields to improve the detection performance of the model. On the other hand, Zhang et al. [15] also proposed surface detection on a specular reflective surface using the ensemble CNN method which integrates two CNN models. These two models are used to extract defective features on the image which then combine toward the end to achieve a better accuracy rate. Based on what has been concluded from past research, it has been decided that the YOLO method is suitable for plastic surface defect detection mostly due to its real-time applicability.

IV. METHODOLOGY

A. *YOLO v5 Architecture*

In this paper, the YOLO v5 model is used to detect and recognize the plastic surface defect. YOLO is a state-of-the-art object detector that enables real-time identification features. In YOLOv5 model, it consists of three main parts: the backbone, neck, and head. The backbone of the model uses the CSP-Darknet (Cross Stage Partial Network). The network performs a CSPnet strategy to partition the feature map of the base layer into two parts and then merges them through a cross-stage hierarchy. The use of a split and merge strategy allows for more gradient flow through the network [16]. The neck section PANet or Path Aggregation Network aims to boost information flow in a proposal-based instance segmentation framework. Specifically, the feature hierarchy is enhanced with accurate localization signals in lower layers by bottom-up path augmentation, which shortens the information path between lower layers and topmost features. Additionally, adaptive feature pooling is employed, which links the feature grid and all feature levels to make useful information in each feature level propagate directly to the following proposal subnetworks. A complementary branch capturing different views for each proposal is created to further improve mask prediction [17]. Finally, the Head is also known as the Yolo Layer that outputs vectors containing class probability, confidence score, and bounding box coordinates [18]. Figure 1 shows the YOLOv5 architecture overview.

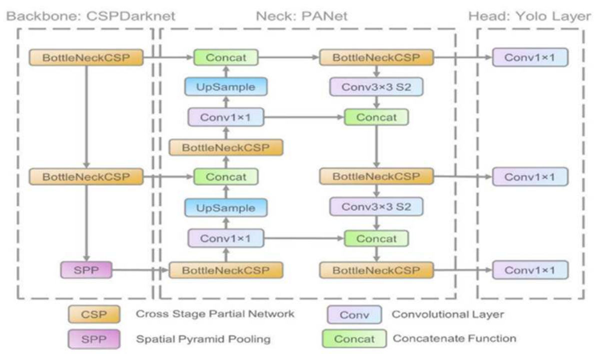

Figure 1. YOLO v5 Architecture Overview [19].

There are four versions of YOLOv5 that consist of small, medium, large, and extra-large sizes. Figure 2 shows the benchmark of each model specification in terms of how well it performs. For this research, the model YOLOv5s is used to train our datasets due to time and hardware constraints.

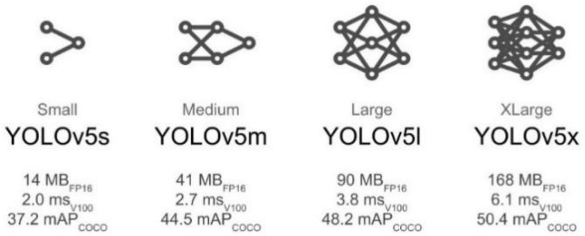

Figure 2. YOLO v5 Models

B. *Performance Evaluation*

To identify the performance of the YOLOv5 model, the precision and the recall rate are used to calculate the mean average precision (mAP) as the network model performance evaluation standard. The mAP is the value of the average detection accuracy of all categories, which is used to evaluate the overall performance of the detection model. The calculation formulas are as follows:

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$mAP = \frac{1}{N}\sum_{i=1}^{N} APi \quad (3)$$

$$F - measure = \frac{2\ x\ recall\ x\ precision}{precision + recall} \quad (4)$$

where TP = True Positive, FP = False Positive, FN= False Negative.

Based on Figure 3, the first step in approaching this research is obtaining the custom dataset. For starters, the defective sample of a 35gm plastic deodorant container has been collected from factory disposal containing various defective surfaces such as black specks, dirt, and scratches. These defective surfaces are later captured as raw images to be classified into a custom dataset. To create a robust dataset, image augmentation techniques such as rotation, flip, shifting, exposure, light density, and color space have been used to increase the quantity and robustness of the raw images. In constructing the YOLOv5 model, Python programming language is used with the aid of the PyTorch library. The YOLOv5 model is constructed according to the YOLOv5 architecture based in Figure 1. Once the training process is completed, the weights are used to evaluate the performance of the model itself in terms of its precision, accuracy, F-measures, and recall.

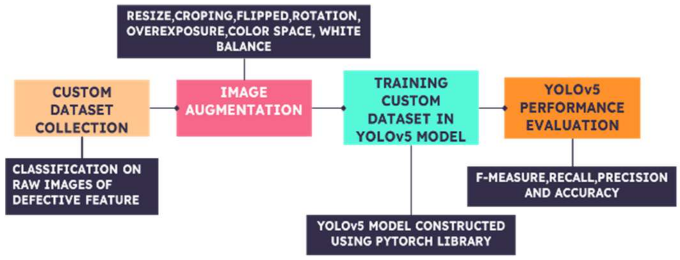

Figure 3. Research Methodology Flowchart

V. EXPERIMENTS

A. *Hardware Using for Detection*

In conducting the experimentation of this research, a Windows operating system equipped with Ryzen 5 CPU @3.6Ghz, NVIDIA RTX2060 graphics card with a 16GB running memory is used, and CUDA 11.0 has been installed to speed up the computation. As for the camera selection in detecting a small object, it requires a high pixelated camera to capture the details on the plastic surface. For that, a CMOS Image Sensor with an 8-megapixel is used. For obtaining the dataset the camera was set to 4K resolution. The raw data used is manually labeled for training and testing purposes. Figure 4 shows the image acquisition platform for obtaining defective features for the custom datasets as well as for the real-time detection test. The setup includes an adjustable stand for the camera to enable a close-up shot of the plastic surface. Moreover, a dedicated light source is also an important component to capture clear images of the defective surface area. The brightness of the light source is adjustable depending on the environmental situation where the exposure rate is required.

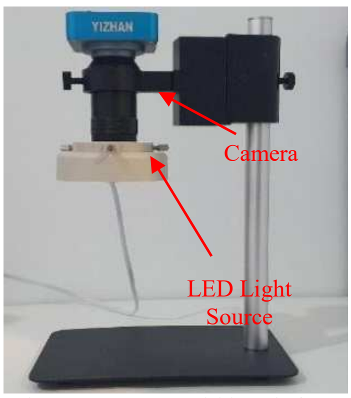

Figure 4. Image Acquisition Platform

B. *Dataset*

In this research, primary datasets used were collected from the plastic packaging factory disposal. The following samples collected were originally from grey pearlescent colored 35gm plastic deodorant containers. Three types of surface defects were able to be identified: black spectacles, dirty surfaces, and scratches. Each defect contains 100 images. To create a robust dataset, each of the defects was enhanced by using image augmentation techniques such as rotating 90 degrees and flipping vertically and horizontally. In addition, color saturation and exposure density parameters were also adjusted to increase the image count to 500 images each. However, due to the shortage of time to set up the experiment, the research will only highlight the identification and classification of black spectacles on the plastic surface container. Table 1 shows some data samples of the defects obtained from the factory disposal.

Table 1. Dataset Sample

Black Spectacles	Dirty	Scratches

C. *Image labeling*

Before training the YOLOv5 model, the custom dataset was manually labeled one by one using the computer vision annotation tool (CVAT). The CVAT software is used to mark the coordination of the location of the defective area and help categorize the types of defects. Figure 5 shows some samples of the labeling process using the CVAT.

Figure 5. Image Labeling

VI. RESULTS AND DISCUSSION

A. *Performance evaluation*

YOLOv5 model training process was conducted by first constructing the network using Jupyter Notebook with PyTorch Library. The performance of YOLOv5 is affected by one of its hyperparameters which is the number of epochs. In the research, three values of epochs are investigated that are 100, 500, and 1000. The training process begins by first entering iterative training for 100 epochs. Figure. 6 shows the performance metrics for the trained network under 100 epochs. Based on Figure 6, we observed that the performance of the network starts to rise around 40 epochs. From the result, we can conclude that the inconsistency in the precision, recall, and mAP is very high which is due to the network being still in the phase of detecting the defects. The smoothing tool was used to process the graph so that we will be able to see the curve much more clearly. From that, it appears that the iterative training for 100 epochs was unable to achieve good stability and consistency in performance metrics. Therefore, to improve the performance of the network, we test run by gradually increasing the iterative training from 500 to 1000 epochs. Based on Figures 7 and 8 we can see how the performance of the network starts to gain more consistency compared to the previous training process.

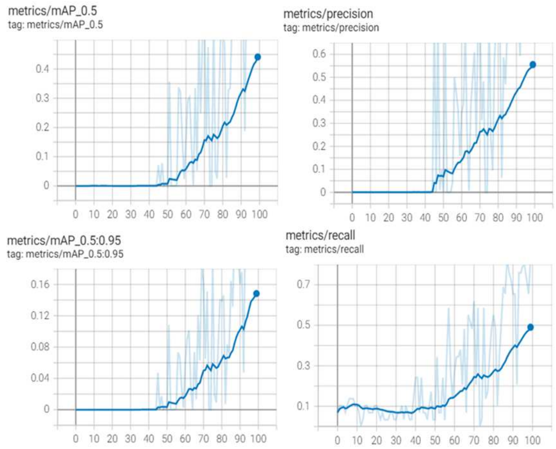

Figure 6. Performance Metrics for 100 epochs.

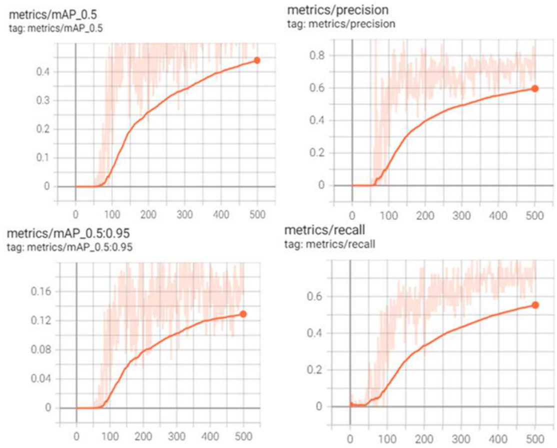

Figure 7. Performance Metrics for 500 epochs.

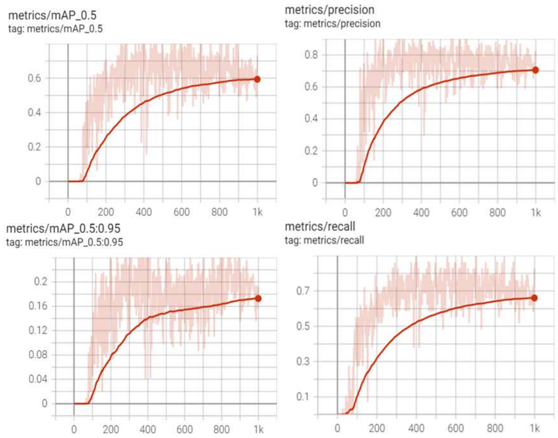

Figure 8. Performance Metrics for 1000 epochs.

From Table 2, we can differentiate between the average precision of the first trained network with 100 epochs and can achieve 67% accuracy. By increasing the iteration value to 500 and 1000 epochs, the detection accuracy has gradually increased by 13% when trained until the 1000 epochs. In addition, the value of its precision, recall, and F-measure were also significantly improved. The time taken for each training process is around 11 minutes for 100 epochs, 58 minutes for 500 epochs, and an hour and 48 minutes for 1000 epochs. Based on the results obtained, YOLOv5 trained with 1000 epochs is used to test the real-time application on a plastic surface defect. Figure 9 shows the real-time testing using the trained YOLO v5 model.

Table 2. Results of YOLO v5 model for 100 epochs, 500 epochs, and 1000 epochs

Defects/AP (%)	100 epoch	500 epoch	1000 epoch
Black Spectacles	67%	68.5%	73%
Precision	0.62	0.65	0.83
Recall	0.58	0.66	0.83
F-Measure	0.6	0.65	0.83

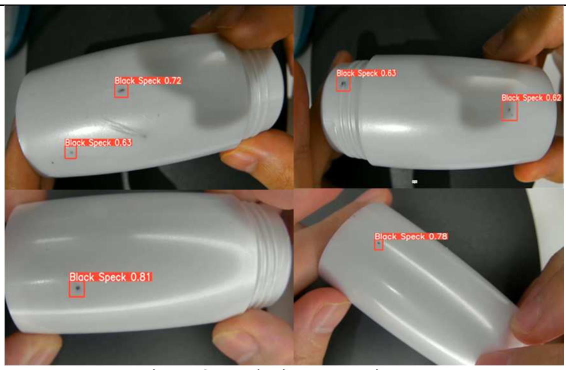

Figure 9. Real-Time Detection Test

VII. CONCLUSION

In this research, real-time plastic surface defect detection based on the YOLOv5 model has been designed and deployed. To ensure that the dataset obtained is high quality, an image acquisition setup was constructed, and image augmentation was performed to create a robust dataset before proceeding to the training process. In training the network, the YOLOv5s model was used, and iterative training is performed for 100, 500, and 1000 epochs to evaluate its performance. It has been concluded that training YOLOv5 with 1000 epochs was able to achieve 73% mAP which is an increase of 13% from the network trained with 100 epochs. For future work, fine-tuning other hyperparameters will be investigated along with the implementation of K-means algorithms to improve the accuracy of the network. Other than that, the implementation of other defects such as scratches and the dirty surface can also be added to determine its performance.

ACKNOWLEDGMENT

This work was supported by the Institute of Graduate Studies of Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia.

REFERENCE

[1] T. N. T. Izhar and Y. V. May, "Life cycle analysis of plastic packaging," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 616, no. 1, 2020, doi: 10.1088/1755-1315/616/1/012036.

[2] H. A. Shari, N. Khalid, N. S. Ashaari, and H. M. Judi, "Statistical process control in plastic packaging manufacturing: A case study," *Proc. 2009 Int. Conf. Electr. Eng. Informatics, ICEEI 2009*, vol. 1, no. August, pp. 199–203, 2009, doi: 10.1109/ICEEI.2009.5254792.

[3] H. I. Lin and F. S. Wibowo, "Image Data Assessment Approach for Deep Learning-Based Metal Surface Defect-Detection Systems," *IEEE Access*, vol. 9, pp. 47621–47638, 2021, doi: 10.1109/ACCESS.2021.3068256.

[4] U. Mittal, S. Srivastava, and P. Chawla, "Review of different techniques for object detection using deep learning," *ACM Int. Conf. Proceeding Ser.*, 2019, doi: 10.1145/3339311.3339357.

[5] Y. W. Chao, S. Vijayanarasimhan, B. Seybold, D. A. Ross, J. Deng, and R. Sukthankar, "Rethinking the Faster R-CNN Architecture for Temporal Action Localization," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 1130–1139, 2018, doi: 10.1109/CVPR.2018.00124.

[6] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *J. Big Data*, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0197-0.

[7] M. An, S. Wang, L. Zheng, and X. Liu, "Fabric defect detection using deep learning: An Improved Faster R-approach," *Proc. - 2020 Int. Conf. Comput. Vision, Image Deep Learn. CVIDL 2020*, no. Cvidl, pp. 319–324, 2020, doi: 10.1109/CVIDL51233.2020.00-78.

[8] Z. Liu, J. Cui, C. Li, S. Ding, and Q. Xu, "Real-time fabric defect detection based on lightweight convolutional neural network," *PervasiveHealth Pervasive Comput. Technol. Healthc.*, pp. 122–127, 2019, doi: 10.1145/3373509.3373550.

[9] Y. Li and Z. Wang, "Research on textile defect detection based on improved cascade R-CNN," *Proc. - 2021 Int. Conf. Artif. Intell. Electromechanical Autom. AIEA 2021*, pp. 43–46, 2021, doi: 10.1109/AIEA53260.2021.00017.

[10] S. Qi, J. Yang, and Z. Zhong, "A Review on Industrial Surface Defect Detection Based on Deep Learning Technology," *ACM Int. Conf. Proceeding Ser.*, pp. 24–30, 2020, doi: 10.1145/3426826.3426832.

[11] X. Tao, D. Zhang, W. Ma, X. Liu, and De Xu, "Automatic metallic surface defect detection and recognition with convolutional neural networks," *Appl. Sci.*, vol. 8, no. 9, pp. 1–15, 2018, doi: 10.3390/app8091575.

[12] Q. Ren, J. Geng, and J. Li, "Slighter Faster R-CNN for real-time detection of steel strip surface defects," *Proc. 2018 Chinese Autom. Congr. CAC 2018*, pp. 2173–2178, 2019, doi: 10.1109/CAC.2018.8623407.

[13] J. Li, Z. Su, J. Geng, and Y. Yin, "Real-time Detection of Steel Strip Surface Defects Based on Improved YOLO Detection Network," *IFAC-PapersOnLine*, vol. 51, no. 21, pp. 76–81, 2018, doi: 10.1016/j.ifacol.2018.09.412.

[14] Z. Zhou, Q. Lu, Z. Wang, and H. Huang, "Detection of micro-defects on irregular reflective surfaces based on improved faster r-cnn," *Sensors (Switzerland)*, vol. 19, no. 22, pp. 1–15, 2019, doi: 10.3390/s19225000.

[15] Z. Zhang, B. Zhang, and T. Akiduki, "Specular reflection surface defects detection by using deep learning," *PervasiveHealth Pervasive Comput. Technol. Healthc.*, no. 3, pp. 6–10, 2019, doi: 10.1145/3325917.3325930.

[16] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," 2020, [Online]. Available: http://arxiv.org/abs/2004.10934.

[17] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path Aggregation Network for Instance Segmentation," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 8759–8768, 2018, doi: 10.1109/CVPR.2018.00913.

[18] E. Onler, "REAL TIME PEST DETECTION USING YOLOv5," no. December 2021, 2022.

[19] R. Xu, H. Lin, K. Lu, L. Cao, and Y. Liu, "A forest fire detection system based on ensemble learning," *Forests*, vol. 12, no. 2, pp. 1–17, 2021, doi: 10.3390/f12020217.