

Assessing the performance of YOLOv5, YOLOv6, and YOLOv7 in road defect detection and classification: a comparative study

Najiha 'Izzaty Mohd Yusof¹, Ali Sophian¹, Hasan Firdaus Mohd Zaki¹, Ali Aryo Bawono², Abd Halim Embong¹, Arselan Ashraf³

¹Department of Mechatronics Engineering, International Islamic University Malaysia, Malaysia

²Faculty of Rail, Transport, and Logistics, Technical University of Munich Asia, Singapore

³Department of Electrical and Computer Engineering, International Islamic University Malaysia, Malaysia

Article Info

Article history:

Received Apr 4, 2023

Revised Jul 14, 2023

Accepted Aug 2, 2023

Keywords:

Machine learning

Object detection

Pavement maintenance

Road crack

Road defect detection

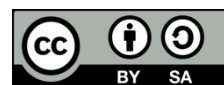
Road inspection

You only look once

ABSTRACT

Road defect inspection is a crucial task in maintaining a good transportation infrastructure as road surface distress can impact user's comfortability, reduce the lifetime of vehicles' parts, and cause road casualties. In recent years, machine learning has been adapted widely in various fields, including object detection, thanks to its superior performance and the availability of high computing power which is generally needed for its model training. Many works have reported using machine-learning-based object detection algorithms to detect defects, such as cracks in buildings and roads. In this work, YOLOv5, YOLOv6 and YOLOv7 models have been implemented and trained using a custom dataset of road cracks and potholes and their performances have been evaluated and compared. Experiments on the dataset show that YOLOv7 has the highest performance with mAP@0.5 score of 79.0% and an inference speed of 0.47 m for 255 test images.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Ali Sophian

Mechatronics Engineering Department, Kulliyah of Engineering

International Islamic University Malaysia, Kuala Lumpur, Malaysia

Email: ali_sophian@iium.edu.my

1. INTRODUCTION

Roads are vital means of transportation in many parts of the world. Various materials are used to construct road pavements, including porous asphalt, stone mastic asphalt and gap graded asphalt, among others. Asphalt is prone to deficiency due to various factors, like being exposed to water and surrounding temperatures, excessive traffic loads, execution mistakes, and lack of maintenance [1]. There are four classifications of different types of defects: pavement cracks, surface deformations, disintegrations, and surface defects. The size and shape properties of road defects can be used to classify them into different categories. They can also be broken down into three severity categories, with mild, moderate, and high severity defects being assessed [2]. Knowledge of the different types of road defects can lead to a better understanding of the probable causes and treatments for defects [3]. As road pavement serves the purpose of having a smooth and comfortable ride and providing surface resistance for safety purposes, any deterioration on its surface must be detected in the early stages for rapid treatment. Road distress identification is also essential to determine the type of maintenance planning needed. There are three categories of detection techniques for road distresses in Malaysia: manual, semi-automatic, and automatic [4]. In recent years, machine learning and machine vision have been adopted in various industry sectors. As it has many benefits in terms of productivity, efficiency, and flexibility with its usage, various fields of study have applied machine learning and machine vision. Despite having various benefit with the developing technologies, some

challenges can also be noticed in the implementation of machine vision in road defects detection, such as hairline cracks that are difficult to be detected, limitation in detecting cracks edge, as well as lack of cracks data quantification for further road maintenance purposes. Recent research on transportation engineering has already explored the application of machine learning technology in detecting road pavement deterioration. convolutional neural network (CNN), artificial neural network (ANN), K-means clustering and regression are some of the most widely used methods thanks to their excellent performance [5].

The main purpose of object detection in road distress inspection is to detect the road defects in the images taken from the inspected roads and correctly classify them according to their types. There are many promising methods of object detection algorithms that are readily available to be adopted. The foremost commonly used approaches are you only look once (YOLO), single-shot detector (SSD) and CNN [6]. CNN is one of deep learning algorithms, which aid in parameter identification by separating image into layers so that each layer is examined and may be interpreted more precisely than the standard analysis approach [7]. Typically, CNN is constructed by incorporating the input, convolutional, pooling, fully connected, and output layers. A network with three convolution layers, two fully connected layers, and two neurons at the output layer since the number of classes needed are for crack and non-crack output [8]. The CNN developed was tested on two different datasets, one obtained from CrackTree200 dataset with an accuracy of 96.99%. At the same time, the another was a self-collected dataset with the highest accuracy of 98.8%. Ma *et al.* [9] tested YOLOv3, YOLOv4s-mish, and YOLOv5s models on timber structures cracks, where YOLOv3 was shown to have the best performance in terms of precision with the mean average precision (mAP) value of 95.5%, while YOLOv5s with mAP value of 92.9% had the fastest training speed because it has the simplest network structure. Meanwhile, Yan and Zhang [10] proposed an algorithm of an improved SSD network by adding a deformable convolution to the backbone feature extraction in detecting asphalt pavement highway crack, resulting with a mAP of 85.11% which is 3.1% higher than the original SSD network.

Horvat *et al.* [11] utilized all of YOLOv5 models to detect face mask in images with a relatively longest training time of 8.67 hours for the YOLOv5x model while having the best performance of 77.1% mAP score. Another YOLOv5 based study introduced by Yu [12], a threshold segmentation method based on Otsu maximum inter-class variance was adopted to the dataset before being trained on YOLOv5-s model. The improved detection achieves 84.37% precision as K-means method has been adapted. Next, Aburaed *et al.* [13] evaluated the performance of YOLOv6 compared to YOLOv5 on detecting craters, where the claims that YOLOv6 would outperform YOLOv5 still can't be proven as their performance was inconsistency in every scenario. Meanwhile, Yang *et al.* [14] proposed a three-stage crack location and segmentation method where it is first filtered by the Retinex method to remove redundant noise, followed by detection process where YOLO-SAMT was introduced, and lastly processed by K-means clustering to extract the cracks. YOLO-SAMT is an enhanced algorithm where YOLOv7 architecture is integrated with SimAM and transformer, which shows a 5.42% higher mAP score than the original YOLOv7. Meanwhile, road damage detection and classification on google street view data using YOLOv7 with a label smoothing technique that resulted in higher F1 scores of 81.7% [15].

The detection and classification of road defects using object detection algorithms such as YOLOv5, YOLOv6, and YOLOv7 face several challenges. Limited availability of high-quality training data, variations in lighting, weather conditions, and road surfaces, and the difficulty in accurately distinguishing between different types of road defects are some of the critical issues to consider. In this context, the objectives of our paper are to evaluate and compare the performance of these algorithms in terms of accuracy, speed, and resource usage, investigate the impact of different data augmentation techniques, explore the use of inference and fine-tuning to improve the accuracy and assess the potential of these algorithms for real-time road defect detection and classification. By addressing these objectives and challenges, this research could contribute to improving the effectiveness and efficiency of road defect detection and classification using object detection algorithms.

This paper is structured into 5 main sections. The section 2 provides an overview of the evolution of the YOLO object detection algorithm, focusing on the YOLOv5, YOLOv6, and YOLOv7 variations. Section 3 outlines the methodology used in this study, including data collection and experimental setup. Section 4 presents the results of the experiments conducted and includes a discussion of these results. Finally, section 5 offers concluding remarks and summarizes the study's key findings.

2. EVOLUTION OF YOLO

YOLO was first introduced in 2015 with the release of "You Only Look Once: Unified, Real-Time Object Detection" paper with main purpose to eliminate multistage of training classifier on bounding boxes and refining them by only executing a single stage of object detection, while ramping up the inference time [16]. Since the release of the first YOLO version, a series of YOLO updated variants has been published by few different scholars with each has its own significant upgrades and features. Following the first version,

published two more papers with the release of YOLOv2 in 2017 and YOLOv3 in 2018 [16]. Bochkovskiy *et al.* [17] continued the variations with the release of YOLOv4 in 2020 as well as YOLOv7. These four versions are established as the official YOLO version, while a lot of other YOLO models such as YOLOR, YOLOX, PP-YOLOE, YOLOv5, and YOLOv6 are labelled unofficial as they are published by other researchers. Among those, a few have more popularity among end users; for example, YOLOv5, published in 2020 by Ultralytics and YOLOv6, released by Meituan Inc in 2021 has comparatively higher performance with its anchor-free method. Few past researches are also published in analysing the performance of YOLO models. Jiang *et al.* [18] compared the differences and relationship of YOLOv1 until YOLOv5 architecture and relativity, where YOLOv4 and YOLOv5 having similar and the highest performance in terms of speed and accuracy at that time. Thuan [19] in his article also concur to the comparison, while expecting more performance value of YOLOv5 as it was newly released at that time. In this paper, the three versions of YOLO; YOLOv5, YOLOv6 and YOLOv7 models, are adapted to compare their performance on road cracks and potholes detection and classification.

YOLO was initially developed to use bounding boxes with a corresponding threshold value to precisely detect objects on images using a model grid cell. YOLOv1 architecture started with the design of Darknet architecture with 24 convolutional layers followed by two fully connected layers inspired by GoogleNet [16]. In the process of improving the algorithm, YOLOv2 was invented with the addition of batch normalization and higher resolution input, as well as replacing the fully connected layers with anchors boxes, which improved the recall by 7% and mAP by 2% [20]. The model is then being developed more with the creation of YOLOv3 with a more powerful backbone, DarkNet-53, with 53 convolutional layers. It eliminates the usage of softmax classifiers, which limits the overlapping boxes, and adopts a logistic regression [21]. Bochkovskiy *et al.* [17] design the enhanced YOLOv4 architecture with the new backbone, combination of cross stage partial network (CSPNet) and Darknet, CSPDarkNet-53, consists of 29 convolutional layers with the addition of spatial pyramid pooling (SPP) block, as well as mosaic data augmentation that uses 4-image mosaic instead of 1 image during training.

2.1. YOLOv5, YOLOv6 and YOLOv7 algorithms

Similar to YOLOv4, YOLOv5 uses CSPDarkNet-53 as its architecture backbone, path aggregation network (PANet) as the neck to improve the effectiveness of data transfer inside the model, and with the addition of a focus layer that replaces the YOLOv3's head layers. However, the developer, Ultralytics, has not released any paper on the model. Even though there are only a few improvements in YOLOv5 architecture compared to YOLOv4, it is the first ever model that implemented PyTorch instead of DarkNet, where PyTorch framework is more user-friendly with language that is widely use in current machine learning technology. Furthermore, with the implementation on the focus technique, YOLOv5 models are 90% smaller than YOLOv4, thus marks a much faster training speed without impacting the mAP score [22]. Figure 1 presents the overall network architecture of YOLOv5 where it consists of three main parts: CSP-Darknet as the backbone, PA-Net as neck, and YOLO layer for the head. CSP-Darknet is a cross stage partial network strategy that is used to help in minimising the excessive amount of duplicate gradient information from usage of residual blocks. This strategy makes YOLOv5 having a faster inference speed due to a smaller number of parameters and computation used. PANet is a feature pyramid network that is utilized in the neck part where it improves in pixels localization. The head of the network for YOLOv5 is similar to YOLOv3 and YOLOv4 where it consists of three convolutional layers that is crucial in calculating the bounding boxes coordinates.

In 2021, Meituan Inc published YOLOv6, designed mainly for industrial applications purposes, also written in PyTorch, is anchor free, and has a reparametrized backbone called EfficientRep where RepVGG is used for nano and small models, while CSPStackRep is used for medium and large models. The neck structure is similar to YOLOv5 with a bi-directional concatenation (BiC) for more localization accuracy, with a decoupled classification and detection head. Overall, YOLOv6 delivers a better result than the former versions in terms of its accuracy and is 51% faster compared to previous anchor-based models [23]. Figure 2 represents the overall network architecture of YOLOv6 [23].

YOLOv7 was released with the publication of the paper, entitled "Trained bag-of-freebies sets new state of the art for real-time object detectors," which revealed a new change of the model architecture by integrating the extended efficient layer aggregation network (E-ELAN) by grouping computational blocks while not changing the transition layers. The architecture is also scaled by concatenating the previous YOLO models for the purpose of inference speed adjustments, as seen in Figure 3. The overall improved architecture of YOLOv7 gives an increasing detection accuracy as well as speed [24].

The overall comparison of the development of YOLO architecture from YOLOv5 up to YOLOv7 can be observed in Table 1. Meanwhile, Figure 4 represents the average precision (AP) curve of YOLO models, where YOLOv7 achieved the highest performance in terms of speed as well as precision [24]. The

following sections present the methodology of implementing the three selected YOLO models; YOLOv5, YOLOv6 and YOLOv7 in detecting and classifying road defects.

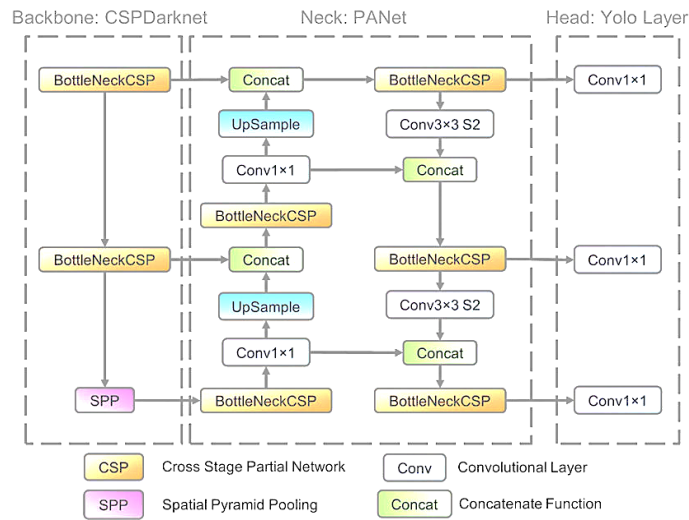


Figure 1. The network architecture of YOLOv5. It consists of three parts: backbone: CSP-darknet, neck: PANet, and head: YOLO layer [25]

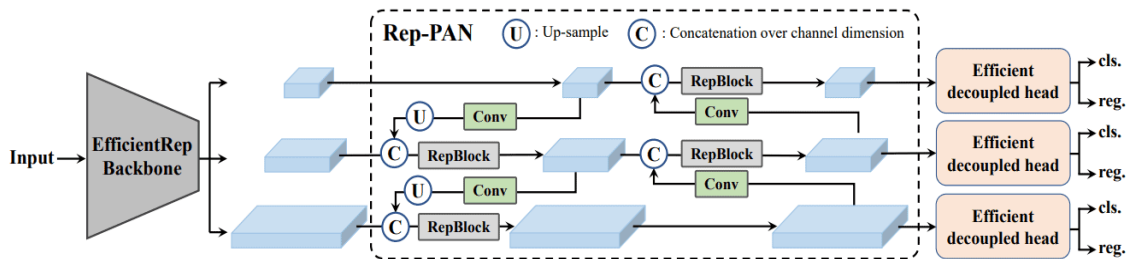


Figure 2. YOLOv6 model architecture [23]

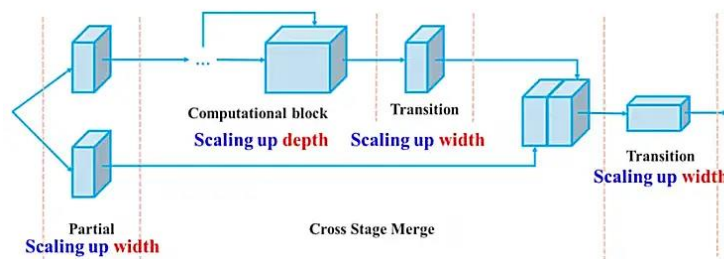


Figure 3. Compound scaling up depth and width for concatenation-based model [24]

Table 1. Architecture structure comparison of YOLOv5, YOLOv6 and YOLOv7

Layers	YOLOv5	YOLOv6	YOLOv7
Backbone	CSPDarknet-53	RepVGG and CSPRepStack	E-ELAN
Neck	PANet	RepPAN	PANet
Head	3 convolutional layers combined with ProtoNet	Decoupled classification and detection head	Lead head and auxiliary head
Loss function	binary cross entropy and logit loss function	VariFocal loss and distribution focal loss	BCE with focal loss and IoU loss

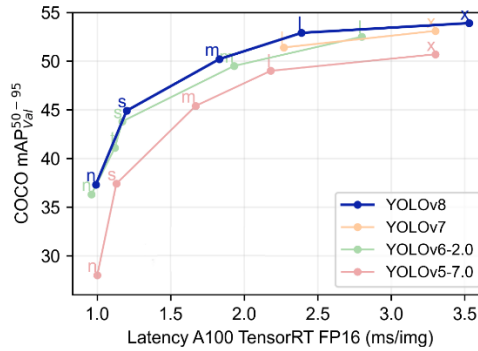


Figure 4. Comparison of YOLO models performance based on AP curve

3. METHODOLOGY

3.1. Data acquisition and pre-processing

The images used in this work were acquired using a GoPro Hero 8 camera mounted behind a car, as illustrated in Figure 5. GoPro Hero 8 offers advantageous features such as image stabilization, lightweight, high-resolution image produced, and practicality. A good image stabilization helps as the camera was mounted on a moving car. GoPro Hero 8 is also practical to be mounted on a car since it is light with only 117g weight and small. Its dimension is 6.2x3.2x4.5 cm. For the data collection, the camera was set to video mode with a 1920x1080 pixels resolution at 24 fps. A linear digital lens was chosen to minimise the barrel effect. The camera was set at a 160 cm height to allow it to capture the road surface at a width of 3.1 m, considered the largest typical width of a road.



Figure 5. Camera setup on vehicle for data acquisition

Videos of the road were captured with format of mp4 for the duration of 5 to 10 minutes at a maximum speed of around 30 km/h. Images were extracted and saved from the videos in jpg format with a resolution of 1920x1080 pixels. A total of 8396 images were extracted from all the videos acquired during data collection, and after manually filtering out images without any visible road defects, 3328 images remained.

Roboflow was chosen as the primary tool to annotate the images, split them, then to augment them. The images were annotated manually using the bounding box features. The annotated defects were split into four classes which are crocodile cracks, longitudinal cracks, transverse cracks, and potholes. The image dataset was then split into train, validation and test sets at the ratio of 7:2:1. The images were then augmented by flipping them in both vertical and horizontal axis resulting in a total dataset of 4788 images split into 4000 training images, 533 validation images and 255 test images, with final image resolution of 640x360 pixels.

The defects to be detected from the images were classified into four classes: crocodile crack, longitudinal crack, transverse crack, and potholes. The sample of the images containing these four classes can be seen in Figure 6(a) for crocodile crack, longitudinal crack as in Figure 6(b), potholes as in Figure 6(c), and lastly transverse crack as in Figure 6(d).



Figure 6. Image samples of road defects captured for each class; (a) crocodile crack, (b) longitudinal crack, (c) pothole, and (d) transverse crack

3.2. Deployment of YOLO models for road crack detection

YOLO models have been chosen because of their proven fast inference speeds and high accuracies. In this work, the performance of YOLOv5, YOLOv6 and YOLOv7 models in road crack detection have been evaluated and compared. The YOLOv5, YOLOv6 and YOLOv7 models were obtained from github.com/ultralytics/yolov5, github.com/meituan/yolov6 and github.com/WongKinYiu/yolov7, respectively. They were trained using the prepared dataset described in the previous section. Google Colab was used for training the models, which offers high-performance GPUs. Roboflow was used to annotate the images, augment selected images, and create the configuration files for model training purposes. The training for each model was completed after 100 epochs. Finally, the inference was also done in Google Colab, although it could have been done locally on a typical laptop and does not require a high processing power. To find the best performing model in terms of both speed and accuracy, many models of YOLO architectures were investigated, which include YOLOv5-n (nano), YOLOv5-s (small), YOLOv5-m (medium), YOLOv5-l (large), YOLOv5-x (extra-large), YOLOv6-n, YOLOv6-s, YOLOv6-m, YOLOv6-l, YOLOv7-tiny, YOLOv7 and YOLOv7-x.

The results obtained from each run were evaluated in terms of precision and accuracy. At the end of each training run, the results were saved, and they include precision, recall, mAP and its mAP at different IoU thresholds ranging from 0.5 to 0.95. The main parameters that need to be focused on are accuracy and mAP@0.5, which is the mean average precision. Meanwhile, as the accuracy result is not included in the data results, it must be calculated using each training run's confusion matrix. The calculations for each of the results are as in (1)-(5):

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

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

$$AP = \frac{1}{11} \sum Precision(Recall) \quad (3)$$

$$mAP = \frac{1}{4} \sum AP \quad (4)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

Where TP is true positive, TN is true negative, FP is false positive, FN is false negative, and AP is average precision.

4. RESULTS AND DISCUSSION

Table 2 shows the performance results of all the models trained in this work. Since all models were deployed using the same instances and dataset for each run, the results can be analysed comparatively. It can be seen that YOLOv7-tiny has the shortest training time despite being in a bigger class range compared to YOLOv5-s and YOLOv6-s. To compare relatively each model to their respective size, YOLOv7 still has an

overall shortest training time among all. Regarding mAP value, YOLOv7 has the highest score of 79.0%. However, YOLOv5-l only lack of 0.01% while having a shorter training time by almost 1-hour difference. YOLOv7 model also records the highest accuracy with 87.16%. Among YOLOv5 models, YOLOv5-l sets the highest performance with 78.9% mAP score and 85.65% accuracy, while for YOLOv6 models, YOLOv6-l take place with a mAP value of 72.32% with a higher accuracy of 86.9%.

Table 2. Training performance results for YOLOv5, YOLOv6 and YOLOv7 models

Model	Training time (hr)	mAP@0.5 (%)	Accuracy (%)
YOLOv5-n	3.71	74.50	86.19
YOLOv6-n	4.47	66.66	84.12
YOLOv7-tiny	2.99	74.80	86.05
YOLOv5-s	4.42	77.00	85.55
YOLOv6-s	5.25	68.11	84.79
YOLOv5-m	4.67	78.40	86.21
YOLOv6-m	6.50	64.86	85.50
YOLOv7	5.78	79.00	87.16
YOLOv5-l	4.92	78.90	85.65
YOLOv6-l	8.33	72.32	86.90
YOLOv5-x	5.84	78.30	85.10
YOLOv7-x	9.25	76.30	86.05

Even though based on the evaluation, YOLOv5-l model has a higher mAP of 78.9% compared to YOLOv5-x, 78.3% as can be seen in Figure 7(a), it can also be observed that YOLOv5-x model has the best performing parameter compared to the other models as it maintains the highest curve throughout the whole run. Meanwhile, from Figure 7(b), YOLOv6-l exhibits the best performance out of the four models. Lastly, YOLOv-7 and YOLOv7-x increase with a similar performance throughout the run while YOLOv-7 outperforms the other on the last few epochs, as shown in Figure 7(c). Figure 8 compares all 3 best models of their respective YOLO algorithms. It can be observed that YOLOv5-l training run has a rapid increase of mAP with the number of epochs in the initial phase compared to YOLOv7, but YOLOv7 outperforms YOLOv5-l towards the final phase of training. Thus, YOLOv5-l, YOLOv6-l and YOLOv7 are chosen as the best models for each respective YOLO version.

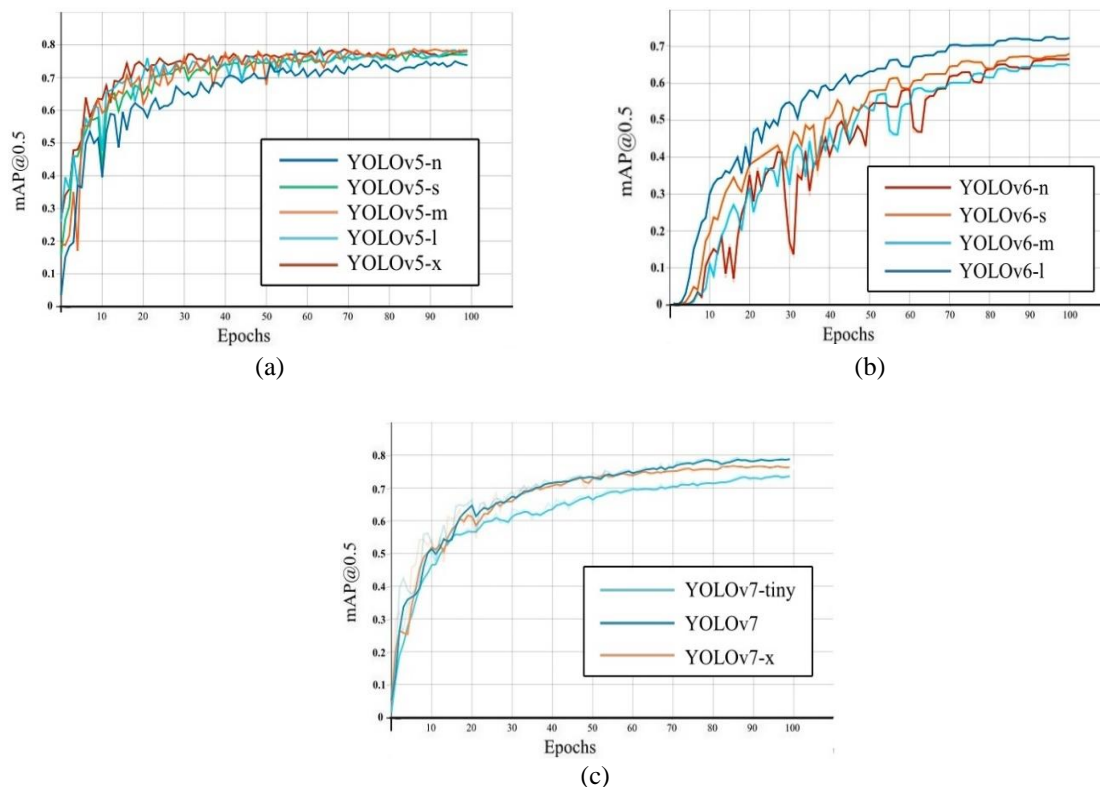


Figure 7. mAP@0.5 curves for models of (a)YOLOv5, (b)YOLOv6, and (c)YOLOv7

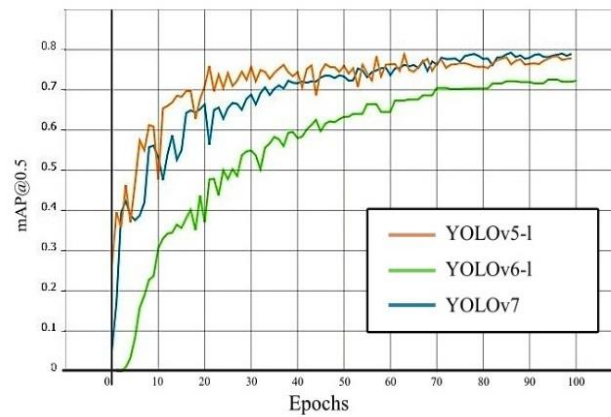


Figure 8. Comparison of mAP@0.5 curves for best models from YOLOv5, YOLOv6, and YOLOv7

To evaluate the results, the best models obtained in the training run, the best models were tested further by inferring other 255 test images to validate the YOLOv5-l, YOLOv6-l and YOLOv7 best models. The speed of the inference run for all best models are recorded in Table 3, with YOLOv7 shown to have the fastest speed.

Table 3. Testing speed for inferring 255 test images using YOLOv5, YOLOv6, and YOLOv7 best model

Model	Inference time (minutes)
YOLOv5-l	0.97
YOLOv6-l	1.68
YOLOv7	0.47

Four sample result images of each best models were compared based on the detection of the crack classes. The confidence score is displayed on the bounding boxes to analyse the models' inference performances, besides the accuracy of identified cracks to their labels. Figures 9 to 12 display the sample inferred images on different type of cracks detected. Figures 9(a) to (c) show the comparison of the confidence score of YOLOv5-l, YOLOv6-l and YOLOv7 in detecting an obvious crocodile crack, where all models give a same high score of 0.98. Figures 10(a) to (c) discussed on the accuracy of detecting multiple cracks on one image and it shows that YOLOv5-l manages to detect the second longitudinal crack that the other 2 models have not detected, as well as having a comparatively higher scores for longitudinal crack and pothole detected. Meanwhile, Figures 11(a) to (c) compares the images with combination of crocodile and transverse cracks which show that the best result is from model YOLOv5-l and YOLOv7 where they have a similar confidence score, with YOLOv5 having a 0.07 score higher in detecting transverse crack. While having a rather lower confidence score in detecting the cracks among all models, YOLOv6-l unexpectedly detected the transverse crack, as shown in Figure 12(b), where the other two models did not detect the obscure cracks at all as seen in Figures 12(a) and (c). From this comparison, it can be concluded that even though YOLOv5-l and YOLOv7 has a very similar performance in inferring the images, YOLOv5 has the upper hand in the confidence score.

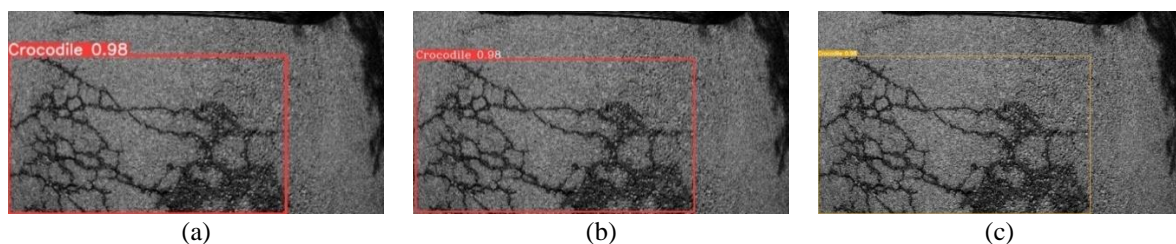


Figure 9. Inference test results on image with a crocodile crack using; (a) YOLOv5-l, (b) YOLOv6-l, and (c) YOLOv7

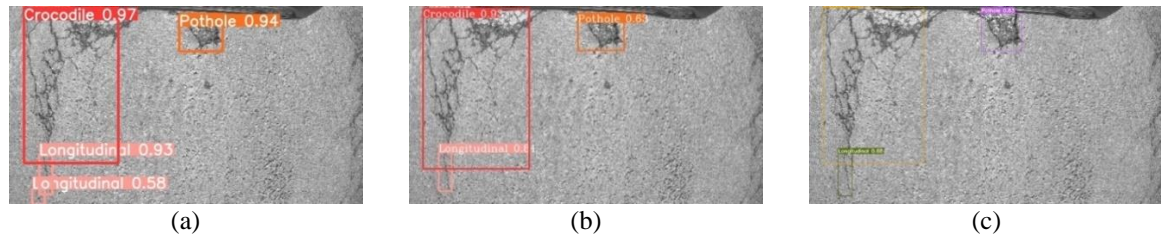


Figure 10. Inference test results on image with a combination of crocodile crack, longitudinal crack and a pothole using; (a) YOLOv5-l, (b) YOLOv6-l, and (c) YOLOv7

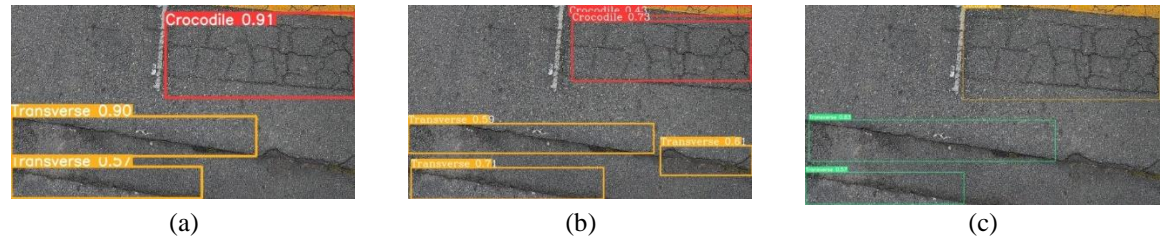


Figure 11. Inference test results on image with a combination of crocodile crack and transverse crack using; (a) YOLOv5-l, (b) YOLOv6-l, and (c) YOLOv7

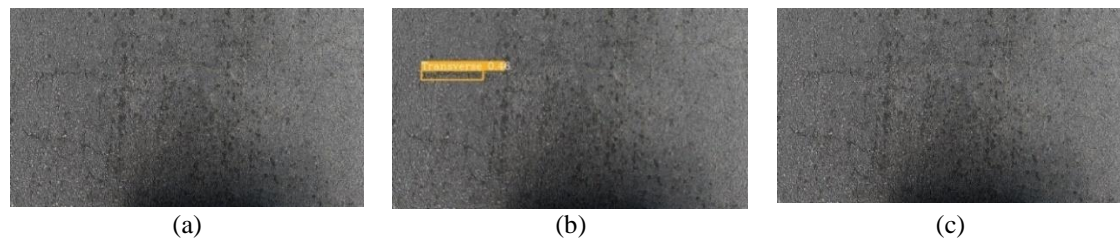


Figure 12. Inference test results on image with an obscure transverse crack using; (a) YOLOv5-l, (b) YOLOv6-l, and (c) YOLOv7

5. CONCLUSION

This paper evaluated the performance of three YOLO models, which are YOLOv5, YOLOv6 and YOLOv7, in detecting and classifying road defects. It was observed that model YOLOv5-l and YOLOv7 have the best implementation among all the 12 models assessed, with a very similar performance. In terms of training execution over a training dataset of 4000 images, YOLOv5 had a training time of 4.92 h, while YOLOv7 trained for 5.7 h, and they evaluated mAP@0.5 score of 78.9% and 79.0% respectively. This shows that YOLOv5 has an upper hand in terms of training performance, as they both resulted a similar precision. In the matter of inferencing process to detect the cracks, YOLOv5 has an inferencing speed of 0.97 minute while YOLOv7 records the speed of 0.47 minute for a total of 255 test images dataset, while they were evaluated with comparison of confidence score where YOLOv5 has higher points. It shows that even though YOLOv7 can perform the inference process at two times faster speed compared to YOLOv5, in terms of accuracy and precision of the detected cracks YOLOv5 still has the advantages. Nonetheless, due to the resource limitations, such as restricting the training run to only 100 epochs and utilizing a dataset comprising only 640 x 360 resolution images and the total images work on was less than 5000, the results were confined to a single discrepancy. To improve upon these findings, future research could entail working on expanded YOLO models and using higher resolution images in conjunction with a variation of epochs number training run. Furthermore, potential pre-processing steps could be implemented on the dataset, and the difference in the dataset inference on images with varying lighting could also be explored.

ACKNOWLEDGEMENT

The authors would like to thank the Malaysian Ministry of Higher Education (MOHE) for financing the research project through the FRGS grant FRGS/1/2021/TK02/UIAM/02/4. We would also like to express




gratitude to the Kulliyah of Engineering, International Islamic University Malaysia for providing the KOE Postgraduate Tuition Fee Waiver Scheme to one of the co-authors.

REFERENCES




- [1] S. S. Adlinge and P. a K. Gupta, "Pavement Deterioration and its Causes," *Mechanical & Civil Engineering*, pp. 9–15, 2009.
- [2] J. S. Miller and W. Y. Bellinger, "Distress Identification Manual for the Long-Term Pavement Performance Program," *Publication of US Department of Transport, Federal Highway Administration*, no. June, p. 129, 2003.
- [3] A. Cubero-Fernandez, F. J. Rodriguez-Lozano, R. Villatoro, J. Olivares, and J. M. Palomares, "Efficient pavement crack detection and classification," *Eurasip Journal on Image and Video Processing*, vol. 2017, no. 39, pp. 1–11, 2017, doi: 10.1186/s13640-017-0187-0.
- [4] N. Hani Mohd Nasir, W. Mazlina Wan Mohamed, K. Nizam Tahar, and S. Alam, "A Review on Road Distress Detection Methods," *Advances in Transportation and Logistics Research*, vol. 1, pp. 230–241, 2018.
- [5] S. R. Karanam, Y. Srinivas, and M. V. Krishna, "Study on image processing using deep learning techniques," *Materials Today: Proceedings*, 2020, doi: 10.1016/j.matpr.2020.09.536.
- [6] A. Duragkar, S. Guhe, A. Sortee, S. Singh, and C. Chandankhede, "Comparison Between YOLOv5 and SSD for Pavement Crack Detection," *ICT Infrastructure and Computing*, vol. 520, pp. 257–263, 2022.
- [7] L. Ali, F. Alnajjar, H. Al Jassmi, M. Gocho, W. Khan, and M. A. Serhani, "Performance Evaluation of Deep CNN-Based Crack Detection and Localization Techniques for Concrete Structures," *sensors*, vol. 21, no. 5, pp. 1–22, 2021.
- [8] M. J. A. Ahmad Faudzi *et al.*, "Detection of Crack on Asphalt Pavement using Deep Convolutional Neural Network," in *Journal of Physics: Conference Series*, 2021, pp. 1–12. doi: 10.1088/1742-6596/1755/1/012048.
- [9] J. Ma, W. Yan, G. Liu, S. Xing, S. Niu, and T. Wei, "Complex Texture Contour Feature Extraction of Cracks in Timber Structures of Ancient Architecture Based on YOLO Algorithm," *Advances in Civil Engineering*, vol. 2022, pp. 1–13, 2022, doi: 10.1155/2022/7879302.
- [10] K. Yan and Z. Zhang, "Automated Asphalt Highway Pavement Crack Detection Based on Deformable Single Shot Multi-Box Detector under a Complex Environment," *IEEE Access*, vol. 9, pp. 150925–150938, 2021, doi: 10.1109/ACCESS.2021.3125703.
- [11] M. Horvat and G. Gledec, "A comparative study of YOLOv5 models performance for image localization and classification," in *Proceedings of the Central European Conference on Information and Intelligent Systems*, 2022, pp. 349–356.
- [12] Z. Yu, "YOLO V5s-based Deep Learning Approach for Concrete Cracks Detection," 2022, vol. 03015, pp. 1–9.
- [13] N. Aburaed, M. Alsaad, S. Al Mansoori, and H. Al-Ahmad, "A Study on the Autonomous Detection of Impact Craters," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, pp. 181–194, 2022, doi: 10.1007/978-3-031-20650-4_15.
- [14] Z. Yang, C. Ni, L. Li, W. Luo, and Y. Qin, "Three-Stage Pavement Crack Localization and Segmentation Algorithm Based on Digital Image Processing and Deep Learning Techniques," *Sensors*, vol. 22, no. 21, pp. 1–31, 2022, doi: 10.3390/s22218459.
- [15] V. Pham, D. Nguyen, and C. Donan, "Road Damage Detection and Classification with YOLOv7," in *Proceedings - 2022 IEEE International Conference on Big Data, Big Data 2022*, 2022, pp. 6416–6423. doi: 10.1109/BigData55660.2022.10020856.
- [16] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016, pp. 779–788. doi: 10.1109/CVPR.2016.91.
- [17] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," *Computer Vision and Pattern Recognition*, 2020.
- [18] P. Jiang, D. Ergu, F. Liu, Y. Cai, and B. Ma, "A Review of Yolo Algorithm Developments," in *Procedia Computer Science*, 2021, vol. 199, pp. 1066–1073. doi: 10.1016/j.procs.2022.01.135.
- [19] D. Thuan, "Evolution of Yolo Algorithm and YOLOv5: The State-of-the-Art Object Detection Algorithm," 2021.
- [20] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017, pp. 6517–6525. doi: 10.1109/CVPR.2017.690.
- [21] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," *Computer Vision and Pattern Recognition*, 2018.
- [22] G. Jocher *et al.*, "ultralytics/yolov5: v7.0 - YOLOv5 SOTA Realtime Instance Segmentation," Nov. 2022. doi: 10.5281/ZENODO.7347926.
- [23] C. Li, L. Li, Y. Geng, and H. Jiang, "YOLOv6 v3.0: A Full-Scale Reloading," *Computer Vision and Pattern Recognition*, 2023, doi: 10.48550/arXiv.2301.05586.
- [24] C. Wang, A. Bochkovskiy, and H. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 7464–7475.
- [25] 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.

BIOGRAPHIES OF AUTHORS






Najiha 'Izzaty Mohd Yusof    graduated in 2018 with a degree in Mechatronics Engineering from International Islamic University Malaysia (IIUM). She was a validation engineer at Intel Technologies for 2 years, specialized in pre-silicone validation mainly working on RTL validation related processes. She also experienced as a Mathematics and Science teacher 2 years before started pursuing a master's degree in Mechatronics Engineering in 2022. Her research interests primarily in the area of machine vision and machine learning. She can be contacted at email:najiha.izzaty.mohd.yusof@gmail.com.






Ali Sophian    is an Associate Professor in the Mechatronics Engineering Department at International Islamic University Malaysia (IIUM). He obtained his B.Eng (Hons.) and PhD, both in Electronics Engineering, from the University of Huddersfield, United Kingdom, in 1998 and 2004 respectively. Prior to joining IIUM in 2014, he used to work in the Mechatronics group of Cummins Turbo Technologies, UK. His research interests include eddy current non-destructive testing, machine vision and machine learning for road inspection applications, and engineering education. He can be contacted at email: ali_sophian@iium.edu.my.






Hasan Firdaus bin Mohd Zaki    is an Associate Professor in the Mechatronics Engineering Department at International Islamic University Malaysia (IIUM) and the Head of Embedded AI at the Centre for Unmanned Technologies (CUTe). He received the Ph.D. degree in computer science from the University of Western Australia. His research interests include robotic vision, RGB-depth object and scene recognition, machine learning, 3D face analysis, and action recognition. He can be contacted at email: hasanzaki@iium.edu.my.






Ali Aryo Bawono    is a Lecturer at Technische Universität München (TUM) Asia in Singapore. He holds a Ph.D. in Civil Engineering from Nanyang Technological University (NTU) in Singapore, a Dr.-Ing. in Civil Engineering from Technische Universität München (TUM) in Germany, an M.Sc. in Transportation System from TUM, and a B.Sc. in Civil Engineering from the Institute of Technology Bandung (ITB) in Indonesia. He worked as Chief Engineer at PT. Jaya Konstruksi Tbk for the toll highway construction project before beginning his research. His research interests include sustainable transport, electromobility and autonomous vehicle mobility, electrified roadway as well as highway and railway design. He can be contacted at email: Ali.Aryo@tum-create.edu.sg.



Abd Halim Embong    is an Assistant Professor in the Mechatronics Engineering Department at International Islamic University Malaysia (IIUM). He earned his master's degree in engineering studies and PhD in Mechanical Engineering (Biomedical) from Auckland University of Technology, Auckland in 2009 and 2015, respectively. Before pursuing his study, he was working as an Industrial Engineering (IE) engineer at Statschippac (M) at Ulu Klang PKFZ for three years. He can be contacted at email: ehalim@iium.edu.my.



Arselan Ashraf    received a Bachelor of Technology (B.Tech.) in Computer Science Engineering from Baba Ghulam Shah University Rajouri, India and a Master of Science (MS) in Computer and Information Engineering from International Islamic University Malaysia. He is currently pursuing Ph.D in Computer and Information Engineering from International Islamic University Malaysia. His research interests include machine learning, signal and image processing, computer vision, and computer network and security. He can be contacted at email: arselanashraf.aa@gmail.com.