Penerbit Akademia Baru

Journal of Advanced Research Design

Journal homepage:
https://akademiabaru.com/submit/index.php/ard
ISSN: 2289-7984

Pothole Detection Using Deep Learning

Izzuddin Hanif Ahmad Husni[1], Mohd Hatta Mohammed Ariff[1,*]

[1] Department of Electronic System Engineering, Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia, 54100 Kuala Lumpur, Malaysia

ARTICLE INFO

Article history:
Received 24 January 2020
Received in revised form 26 June 2020
Accepted 9 July 2020
Available online 22 July 2020

Keywords:
Pothole; autonomous vehicle; YOLOv4; convolution neural network

ABSTRACT

Malaysia is a country that experience hot and rain conditions throughout the year. Due to this climate situation, potholes can be easily formed on the road surface. The formation of potholes happens when there is a crack on the pavement surface allowing rain water to enter through the crack. The crack is formed due to the expenditure of asphalt during hot periods and the contraction of asphalt during cold temperatures. To avoid the damage and injuries caused by potholes to vehicles and drivers, the idea of pothole detection was created. The potholes detection was estimated to help drivers avoid potholes by giving an early warning. The detection can also help Autonomous Vehicle (AV) industry to create a vehicle that can avoid potholes automatically. The detection method that was used in this project was a 2D vision-based method that has low development cost due to applying camera as the detection device. The 2D vision-based method was in computer vision field, combining with deep learning. You Only Look Once Version 4 (YOLOv4) was used as object detection algorithm due to fast detection in real-time and high Average Precision (AP) value. The selected object detection algorithm used Convolution Neural Network (CNN) architecture, containing 75 convolution layers. Image annotation tool software named labelimg was used to create the dataset for training an object detection. The finished potholes detection procedure was tested with the same image but different light quantity. The test was separated into three different light exposer, high, low and very low. The purpose of this test was to measure the effectiveness of the detection during high light quantity, low light quantity and very low light exposures. The parameters were number of batches iteration and threshold value. From the analysis 9000 batches iteration and 90% threshold value was chosen to run the pothole detection. Based on these two selected parameters used during the detection process the project was a success with Average Precision (AP) above 80% and frame rate as low as 30 FPS. Overall, this method was the effective way to increase the detection accuracy and confidence.

1. Introduction

Pothole is one of the common problems of the road. It can damage the car rims or burst the tires. The worst case by hitting pothole was a road accident that cause death. Vehicles such as motorcycle and bicycle also receive the damage caused by potholes. Motorcycle users has the highest risk of death from hitting potholes because of less protective gear. The situation can be

* *Corresponding author.*
E-mail address: mohdhatta.kl@utm.my

much dangerous in wet conditions, with slippery road surface causing less tyre grip and potholes fill with water, make the probability of accidents to occur.

Most potholes are formed due to stress on the road surface. The chucks of pavement between cracks are loose and may eventually be picked out of the surface by continued wheel loads, thus forming a pothole [1]. The formation of potholes in Malaysia usually caused by water entering through the layer of asphalt. The characteristic of the asphalt is that it will become soft when hot and hard when it cools, due to the climate in Malaysia's hot and rain condition throughout the year, the formation of potholes will be more often. Potholes are also formed due to soil deposition at the bottom layer of asphalt. This problem occurs because there is air pocket in the ground, under the surface of the road pavement. This situation occurs due to the process of accelerated road construction without analysis of the construction site geography.

Since potholes are easily formed in Malaysia's climate, a lot of accident have been reported caused by potholes. Nowadays road accident in Malaysia caused by potholes has been a major contributor. The most attracted pothole accident on media in Malaysia was an accident involving Malaysia's Minister of Science and Technology [2], on 27 Dec 2020, during cycling. The accident caused several injuries on his face and body.

A research by Darma *et al.*, [3], reported that the highest proportion of deaths due to road defects (48.6%) is associated with lack of street lighting provision (15.4%) and that potholes (11.2%) contribute to road traffic deaths [3]. Potholes will be more fatal during night and rainy day, during the night potholes are hardly detected making it difficult to avoid.

With the expanding on knowledge and technology in near future, an Autonomous Driving System (ADS) is no longer impossible. To create a much safer vehicle with ADS, real-time detection of pothole can be a crucial factor for autonomous vehicle system. Since potholes can be the major cause of accident for this system and disturb the autonomous vehicle motion strategy, the shock of collision between potholes can damage the system and sensor on the vehicle. This problem can cause traumatic accident and big hurdle to the Autonomous Vehicle (AV) Company. Therefore, pothole detection is considered crucial to ensure the safety of an AV. The aim of this study therefore was to develop a pothole detection method with the best object detection algorithm that can be used in real-time application.

2. Detection Methods

2.1 Vibration-based Methods

Vibration based methods are commonly used for structural health monitoring (SHM) and damage detection (DD). The simplest vibration-based methods are the methods based on the model analysis of the structure [4]. The natural frequencies are easily measurable with high accuracy and are usually less contaminated by experimental noise. To detect the potholes on the road surface, this method uses accelerometer to record the vibration produced by the pavement surface. The responses from the vibration will estimate the road condition and the depth of potholes. However, the potholes in the middle of the road have a difficulty to be detected using this method due to the accelerometer being places on the vehicle wheels and most of the time the car wheels will be on the side of road. Sometimes the result could be wrong as there are a few joints on the surface of a road such as bridge and sewerage steel cover. Some of the researches that were done using this method involved a survey on pavement condition proposed by B. X. Yu and X. Yu [5]. Another project that utilized this method was proposed by De Zoysa *et al.,* [6] and Eriksson *et al.,* [7] both of them proposing the same method by combining accelerometers and GPS, to detect and locate the position of potholes at the same time. The project proposed by Rode *et al.,* [8] however, integrated pothole detection via vibration-based method and warning system. The project was considered helpful to the drivers who were unaware of potholes during driving. Many of the proposed projects and researches use this method due to low development cost and small data size can be stored.

2.2 3D Reconstruction-based Methods

The 3D reconstruction-based method is a method that recreate 3D model using equipment such as 3D laser scanner (LIDAR), multiple camera and visualization using depth sensor. The detection using 3D laser scanner method will provide much detail than other methods such as distance, position and depth. Many researchers applied this method, including Chang *et al.,* [9] who produced an accuracy error by a millimetre using time-off light scanners that managed to extract specific distress on pavement surface. Another researcher managed to develop an inspection system to differentiate distress on pavement surface such as potholes. Li *et al.,* [10] successfully created 3D geometric features with vertical and horizontal resolutions of 2 mm. Both researchers stated that 3D laser scanner has an accurate value in real-time. However, this method required a lot of development cost due to the price for 3D laser scanner is expensive.

The 3D reconstruction-based method using multiple cameras has much lower cost compared to 3D laser scanner. Wang [11] used this method with two digital cameras to detect pavement distress and classified it. The method applied same images from different cameras angle to reconstruct a 3D model, however the method is not accurate as using 3D laser scanner. Hou *et al.,* [12] applied the same method with four digital cameras. The hardest part of using multiple cameras was it needed high computational efforts to recreate the pavement surface in 3D model.

Depth sensor is another equipment that is categorized as a 3D reconstruction-based method. Depth sensor has the lowest development cost in 3D reconstruction-based method. Joubert *et al.,* [13] has developed a pothole detection and analyser with GPS coordinates, applying a depth sensor and high-speed USB camera. Another researcher, Moazzam *et al.,* [14] used depth sensor to generate 3D pothole image and calculate the geometrical information. From that, the pothole area, volume and depth can be estimated. Mahmoudzadeh *et al.,* [15] has proposed the best approach for pavement management was through using depth sensor which is precise and cost-effective.

2.3 2D Vision-based Methods

2D vision-based method so far is the best method for object detection. In term of potholes detection, it has the balance between the 3D reconstruction-based methods and vibration-based methods. Since 2D vision-based method require camera as detection equipment, the development cost will be low. There also a lot of open-source object detection algorithm that can be used. It has high accuracy depending on object detection algorithm and light intensity. There are two types of 2D vision-based methods, first is 2D image-based approaches and second, 2D video-based approaches. Based on 2D vision-based method Lin and Liu [16] successfully created a pothole detection by extracting the measured texture based on histogram as the image ground truth region. They used the nonlinear support vector machine to identify the image ground truth region. The method that was proposed by Koch and Brilakis [17] used the positive and negative region of ground truth image and called it as defect and non-defect region. The image that is potentially deemed a pothole will be analysed, the texture of pothole will be extracted and labelled as defect shape. The surrounding area of pothole (pavement) will be marked as non-defect region. The compared input image with defect and non-defect will determine the region that represent potholes. By using the image processing spectral clustering method, Buza *et al.,* [18] has managed to develop a pothole detector. Another 2D vision-based method proposed by Ryu *et al.,* [19] which is the method that combined image processing and geometric information is to extract image ground truth. The defect region will be compared with non-defect region in term of standard deviation and histogram similarity to identify the potholes.

For video-based approaches, a few researchers have used this method in 2D vision-based method. By combining the standard deviation, area, roundness and diameter information, Huidrom *et al.,* [20] managed to detect and classify pothole in video clips via heuristic decision logic. With vision tracker to trace detected potholes, Koch *et al.,* [21] accomplished to incrementally update the texture signature for the pavement region. This method can evaluate the condition of potholes. Tedeschi and Benedetto [22] detected potholes by applying computational resources provided by an Android-based device. The method used the recognition system in OpenCV library.

2.4 Deep Learning and Object Detection Algorithms
2.4.1 YOLOv4: Optimal speed and accuracy of object detection

YOLOv4 [23] was created by a group of researchers led by Alexey Bochkovskiy and two researchers from the Institute of Information Science Academia Sinica, Taiwan namely Chien-Tao Wang and Hong-Yuan Mark Liao. The approach that YOLO applied [24] for object detection was, to frame object detection as a regression problem to spatially separate bounding boxes and associate class probabilities. YOLO used single neural network to predict bounding boxes and class probabilities from a full image in any one evaluation. Also since YOLO used a single network pipeline, the detection performance can be optimized end-to-end directly. The major different of YOLOv4 and YOLOv3 [25] was their backbone, the YOLOv3 used Darknet53 method as its backbone while YOLOv4 used CSPDarknet53 method. YOLOv4 combining CSPNet [26] backbone with Darknet53 methods created CSPDarknet53 method. Currently YOLOv4 is the fastest and most accurate in real-time object detection.

2.4.2 Learning rich features at high-speed for single-shot object detection

Learning Rich Features at High-Speed for Single-Shot Object Detection was created by a group of researchers that included Tiancai Wang from School of Electrical and Information Engineering, Tianjin University and Rao Muhammad Anwer, Hisham Cholakkal, Fahad Shahbaz Kan, Yanwei Pang and Ling Shao from Inception Institute of Artificial Intelligence (IIAI), UAE. The objective of this research was to reduce the training time of detection model trained from scratch, compared to pre-trained models and maintaining the benefit to reduce the task gap between classification and localization, especially at high overlap thresholds. To achieve their objective, they introduced a single-stage detection framework that combined the advantages of pre-trained models and training from scratch. The framework that they applied contained pre-trained backbone and a parallel light-weight auxiliary network trained from scratch [27].

2.4.3 Receptive field block net for accurate and fast object detection

Receptive field block net for accurate and fast object detection was created by a group of researchers from Beijing Advanced Innovation Center for Big Data and Brain Computing, Beihang University. The study was conducted by Liu *et al.*, [28]. The objective of their research was to explore an alternative to build a fast and accurate detector by strengthening lightweight features using a hand-crafted mechanism. The research was inspired by the structure of Receptive Fields (RFs) in human visual systems. They created a Receptive Fields Block (RFB) module, which imitates receptive fields of human. They, then combined the RFB to the top of Single-Shot Object Detection (SSD), creating RFB Net detector [28].

3. Methodology
3.1 Training Object Detection Work Flow

Figure 1 shows the process of training object detection involving 2 phases. Phases 1 and 2 are the preparation and training phases, respectively. In phase 1, image annotation was the first step for training dataset where all the ground truth in the dataset being labelled according to their class. The next step in phase 1 was compiling all files needed for training in a folder. The files include dataset, configuration file, data file and pre-trained weight.

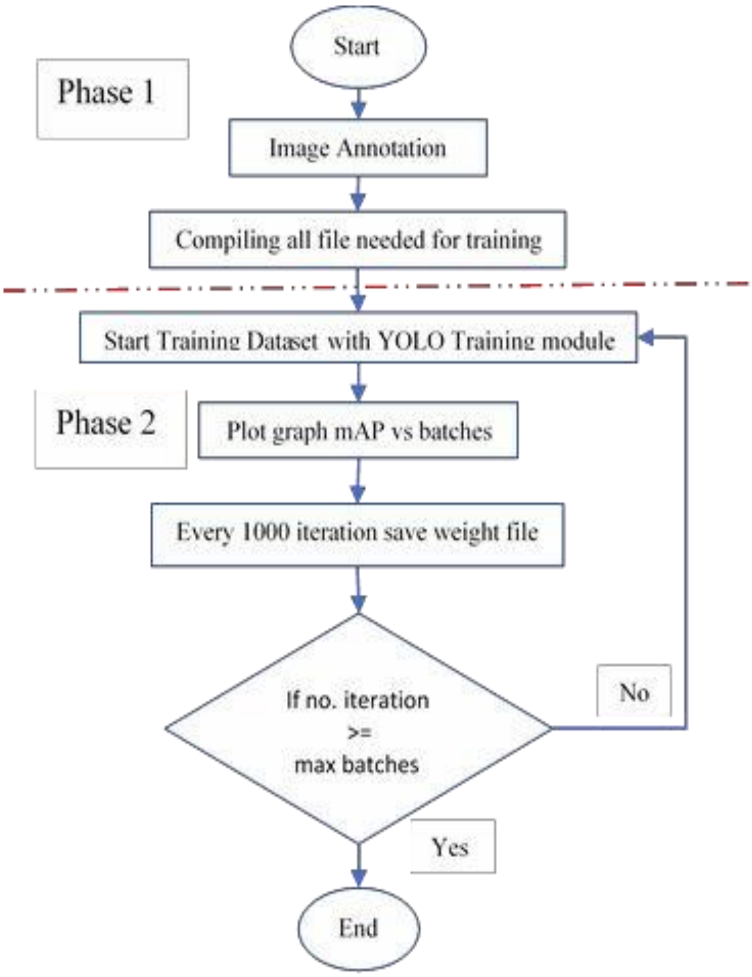

Fig. 1. Process of object detection training

The training phase started in Phase 2. The training dataset starts with YOLO training module, where the code to start the training in darknet framework was executed. When the code completes compiling the graph mean Average Precision (mAP), then the plotting exercise of Average Loss versus Batches will start. Then for every 1000 iteration of batch, the weight file will be saved in backup folder. The number of iterations then, will be compared with the maximum batches. If the number of iterations is equal to the maximum batches the training will stop, if the number of iterations is less than maximum batches, the training will continue.

Figure 2 shows the process of using trained weight for object detection. The process involves two phases. Phase 1 is the preparation phase and phase 2 is the training phase. In phase 1 all the files needed to start the training will be compiled in one folder. The files needed were data, configuration and object detection trained weight files.

For phase 2 where the object detection started in the darknet framework, the input image will be read by YOLO testing module. When the YOLO testing module code started, it will start compiling all the files and weights. It will analyse the input image finding the pattern of ground truth. If the testing module find the ground truth in input image it will display a rectangular marker on the ground truth and it class, if not it will end the detection.

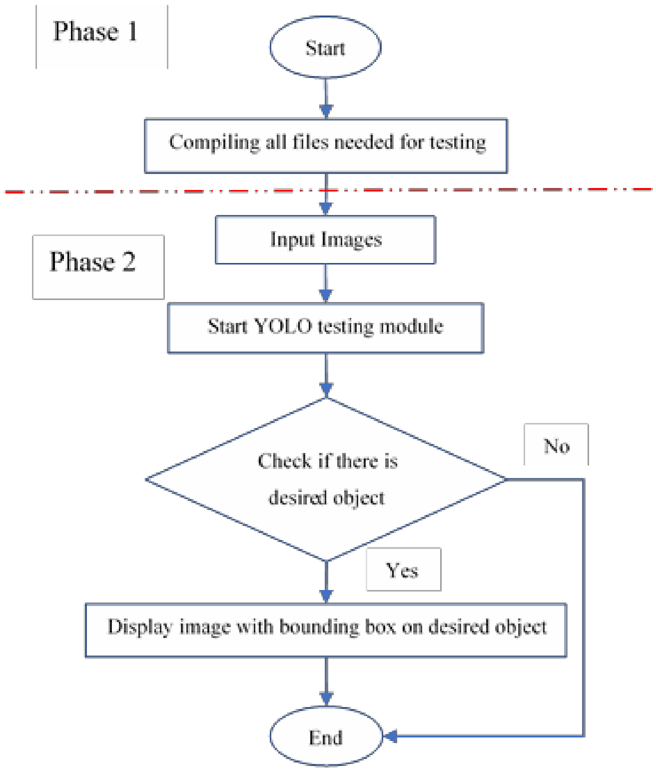

Fig. 2. Process of detecting object using trained weight

3.2 Selecting Weight File

After completing the whole training process, as the result there will be a few weight files created according to the maximum batches. There were several factors to decide which weight file has the precise detection and high accuracy. The first factor was value of average loss, the lowest value of average loss will give more detection accuracy. The second factor was the percentage of mAP. Here, as the obtained mAP value gets higher, it will give the detection system higher confident level.

To select the best weight file, an analysis was conducted using pothole image with three different conditions of light exposer. Figure 3-5 was the pothole image that been used during the analysis to decide the best weight file for the project. There were two potholes in each figure labelled A and B. The first picture will have high light exposer followed by low light exposer and very low light exposer. The weight file that has the highest confident level in these three conditions was selected as the best weight file for the project.

Fig. 3. Pothole with high light exposer

Fig. 4. Pothole with low light exposer

Fig. 5. Pothole with very low light exposer

4. Result and Discussion

4.1 Result from Training Dataset

The training process was completed after it reach 9000 batches iteration. Every 1000 batches iteration the program was saved in form of weight files. Each iteration has mAP of detection obtained as guide for selecting the best weight files. The obtained mAP during the training has Intersection over Union (IoU) threshold of 50%. Table 1 show the result of mAP percentage for every one thousand iterations of batches. Analysis result on threshold values used same value of batches iteration to test every possible iteration.

Table 1
mAP percentage for 50% threshold

No. Iteration (Batches)	mAP Percentage IoU threshold 50 %
1000	66.13%
2000	72.15%
3000	76.83%
4000	73.69%
5000	76.04%
6000	77.17%
7000	71.88%
8000	74.38%
9000	75.34%

Graph in Figure 6 was plotted based on data in Table 1. The mAP percentage increase when the number of batches iteration increase. The increase value of iteration number also will cause the Average Loss (AL) to decrease. There was sudden drop of mAP percentage from iteration 3000 batches to 4000 batches and from iteration 6000 batches to 7000 batches, the pattern can be seen in Figure 6. This sudden drop was cause by pausing and resuming the training process. This makes mAP percentage not a good reference for selecting the best weight file. The training stops at 9000 iterations due to limitation of hardware, adding more iteration will take much longer time with little improvement of detection.

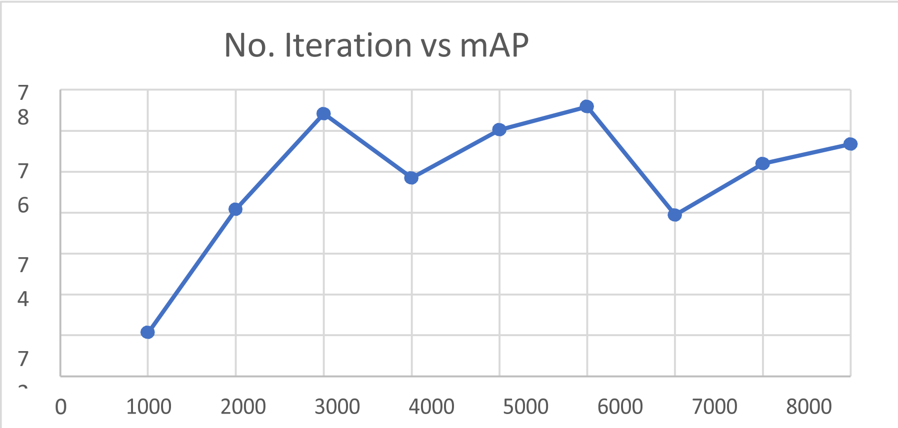

Fig. 6. Number of Iteration verses mAP Percentage graph

4.2 Analysis Result on Weight File Iteration

Using the pothole image in Figure 3-5, the image with difference light exposer was used to test the confidence level. Since in the images has two pothole it was labelled as A and B. The constant value of threshold was set to 0.25 during the test. Table 2 show the potholes confidence level for every one thousand iterations with difference level of light exposer.

Table 2
Percentage of confidence level with difference light exposer

No. Iteration (Batches)	Percentage of confidence level					
	High light exposer		Low light exposer		Very low light exposer	
	A	B	A	B	A	B
1000	0.72	0.56	0.60	0.31	0.55	0.00
2000	0.95	0.77	0.95	0.80	0.82	0.59
3000	0.94	0.89	0.93	0.78	0.71	0.57
4000	0.97	0.87	0.96	0.94	0.00	0.92
5000	0.99	0.91	0.99	0.96	0.80	0.97
6000	0.99	0.90	0.99	0.93	0.79	0.95
7000	0.98	0.97	0.98	0.88	0.81	0.92
8000	1.00	0.96	1.00	0.98	0.93	0.86
9000	1.00	0.98	1.00	0.98	0.94	0.83

From Table 2 the confidence level of pothole detection increases when the number of iteration increase. At iteration 9000 batches the detection has the highest confidence level in high light exposer. In low light exposer there was a few numbers of iteration that has higher confidence level than high light exposer such as, from iteration 4000 batches to 6000 batches, pothole B has higher confidence level than high light exposer due to the training dataset has much darker images than bright images.

Considering a very low light exposer, the 9000 batches still have the highest confidence level for pothole A and 5000 batches have the highest confidence level for pothole B. The unpredictable value of confidence level due to the data was never been train and the ground truth image start to loss its pothole pattern.

4.3 Analysis Result on Threshold Values

The false detection occur due to the other object has some similar pattern with the trained object. To overcome this problem a test on road video was conducted to fine the suitable threshold value set for the pothole detection. Table 3 show the test result based on input threshold value. The 7000 iteration batches weight file was used the test as a constant.

Threshold value 0.25 was the default value for YOLOv4 Object Detection. Based on result in Table 3 the false detection made with 0.25 threshold has 27 false detection, 9 pothole detected and 0 miss detection. The detection on threshold value 0.25 not only detect the manhole, puddle and defective road lane, it also detects dirt on front windscreen. So, the threshold value below 0.25 was not valid.

From Table 3, threshold value 0.85 has reduce the false detections from 27 to 7, 2 miss detections and 7 potholes detected. The miss detection of the pothole was not very crucial for driving car because the pothole was on the other road side. The threshold values below 0.85 was not included in the table because it has a lot of false detections. Starting from this threshold the false detection reduced to only puddle and manhole. The increasing of threshold value makes the number of false detections reduced and the number of true detections also decrease.

Table 3
Output for threshold value

Threshold	Pothole Detected		False Detection	Miss Detection
	Driving Lane	Outside Driving		
0.25	5	4	27	0
❙	❙	❙	❙	❙
0.85	5	2	7	2
0.86	5	2	4	2
0.87	5	2	4	2
0.88	5	2	3	2
0.89	5	2	3	2
0.90	5	2	2	2
0.91	5	1	1	3
0.92	5	1	1	3
0.93	5	1	0	3
0.94	5	1	0	3
0.95	5	0	0	4
0.96	5	0	0	4
0.97	5	0	0	4
0.98	5	0	0	4
0.99	2	0	0	7
1.00	0	0	0	9

4.4 Selecting Weight File and Threshold for Object Detection

Based on Table 4, weight file with iteration 9000 batches was chosen to perform the object detection because it has the highest confidence level. Based on Table 4, value 0.90 was selected as threshold for the object detection because the number of false detections was 2 and 2 miss detections. Table 4 show the mAP percentage for every iteration with default threshold value 25 % and selected threshold value 90 %.

Table 4
mAP percentage for 25 % and 90 % threshold

No. Iteration (Batches)	mAP Percentage	
	IoU threshold 25 %	IoU threshold 90 %
1000	76.63 %	0.03 %
2000	81.78 %	0.08 %
3000	81.85 %	1.13 %
4000	81.28 %	1.24 %
5000	81.99 %	3.07 %
6000	83.03 %	1.31 %
7000	79.12 %	0.29 %
8000	82.02 %	3.74 %
9000	81.71 %	5.69 %

The reason of not choosing threshold value 0.95 was due to 4 miss detections compare to threshold value 0.90 that has 2 miss detections. Table 4 show that higher threshold will give lower mAP. So, by increasing higher threshold value has the higher risk of miss detecting smaller pothole and slow the process of detection.

Images in Figure 7 show the differences between threshold 25 % and 90 %. In threshold 25 % image, it shows that the AI also detecting puddle, manhole and defection lane. By applying much higher threshold such as images with 90 % threshold value to the detection parameter, the false detection was no longer occur.

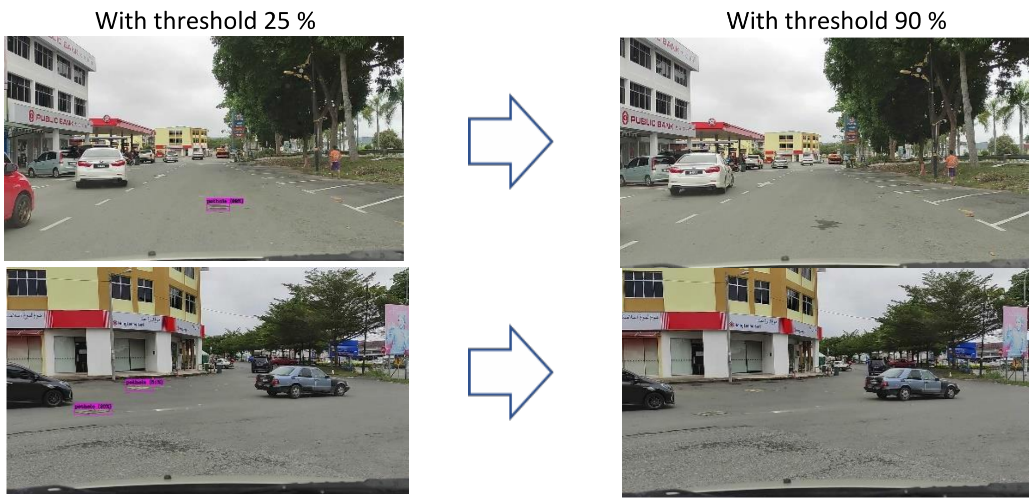

Fig. 7. Differences between threshold 25 % and 90 %

4.5 Result for Object Detection

The result for this project as Figures 8-10 was using the data gather in Table 4 which is 9000 batches iteration weight file and 0.90 threshold value. The pothole will be coordinated with a bounding box. The computer would able to recognize the image of pothole. As in real-time application the prediction would be at least 30 FPS or greater. Since this project using 2D-Vision Based Method, it will not provide distance of the pothole. The only parameter that can get is the coordinate of bounding box edges.

The image in Figure 8 was a result from high light exposer detection test in Figure 3. From the image in Figure 8, both potholes were detected using the object detection algorithm. Pothole A has detection confidence level of 100% while pothole B 96%. The result in Figure 8 has no significant difference from other test that run at 0.25 threshold value in Table 2 with only 2% different on pothole B.

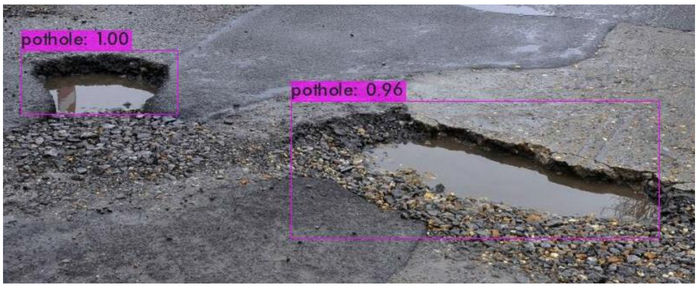

Fig. 8. High light exposer detection image

Figure 9 show the result of low light exposer detection in Figure 4. Both potholes in Figure 9 were able to detect using the detection algorithm. Pothole A has detection confidence level of 100% and pothole B 98%. The result of pothole B has slightly higher confidence level in low light exposer compare to high light exposer was expected based on result in Table 2. The factor of this was due to there was a lower light exposer than high light exposer.

Fig. 9. Low light exposer detection image

Very low light exposer result in Figure 10 was based on Figure 5. In Figure 10 only pothole A was able to detect using the algorithm with confidence level 93%. The high value of threshold makes the detection for pothole B difficult for the algorithm to detect. Hance the image in Figure 10 has very low light exposer make it hard to see even for human vision. The detection algorithm can detect the pothole when the threshold value reduced.

Fig. 10. Very low light exposer detection image

Fig. 11. Very low light exposer detection image, 50 % threshold value

Figure 11 show the detection result of very low light exposer image with threshold value 0.50. The algorithm able to detect both potholes with 94 % confidence level on pothole A and 83 % on pothole B. The main reason was to eliminate the false detection of the algorithm.

4.6 Identified Suitable Threshold

Sometimes the detection system makes a false detection due to the object have some similarity to potholes. This false detection can be object such as manholes, defective road lines and puddle. By setting the right value of threshold, the false detection can be eliminated or minimized their confidence level.

By default, the value of threshold was 0.25 and the range value for threshold in YOLOv4 object detection was from 0.00 to 1.00. To identify the suitable value of threshold for this project a car dashboard recorded video was used. In the video there was total of eleven potholes. The threshold value that has the fewest false detection was identified as the threshold value for the project.

5. Conclusions

The detection algorithm was able to make the detection during high and low light exposer, but has a difficulty to make the detection during very low light exposer due to lack of light. The second objective, to analyses the performance of pothole detection accuracy and precision based on number of training iteration and threshold value, also contribute to the success of the project. Without proper analysis on the pothole detection performance some of the false detection can occur. So, this pothole detection can be an effective solution in order to increase the safety of Autonomous Vehicle and others vehicle.

From the result, the project of developing pothole detection using deep learning was success. By using YOLOv4 as object detection algorithm it was proven to have the accurate and precision detection with high FPS among others object detection algorithm. In the result with parameters 9000 batches iteration and 90 % threshold value the pothole detection was able to detect pothole and minimize the false detection such as manhole, puddle and defection road lanes.

References

[1] Sharma, Umesh, A. Kanoungo, Atal Nagar, and Himachal Pradesh Barotiwala. "Study of causes of potholes on bituminous roads–a case study." *Journal of Civil Engineering and Environment Technology* (2015): 345-349.

[2] The Straits Times. Malaysia’s Minister Khairy Jamaluddin injures from fall after bicycle hits pothole, 28 December 2020.

[3] Darma, Yusria, Mohamed Rehan Karim, and Sulaiman Abdullah. "An analysis of Malaysia road traffic death distribution by road environment." *Sādhanā* 42 (2017): 1605-1615. https://doi.org/10.1007/s12046-017-0694-9

[4] Manoach, E., J. Warminski, L. Kloda, and A. Teter. "Vibration based methods for damage detection in structures." In *MATEC Web of Conferences*, vol. 83, p. 05007. EDP Sciences, 2016. https://doi.org/10.1051/matecconf/20168305007

[5] Yu, Bill X., and Xinbao Yu. "Vibration-based system for pavement condition evaluation." In *Applications of advanced technology in transportation*, pp. 183-189. 2006. https://doi.org/10.1061/40799(213)31

[6] De Zoysa, Kasun, Chamath Keppitiyagama, Gihan P. Seneviratne, and W. W. A. T. Shihan. "A public transport system based sensor network for road surface condition monitoring." In *Proceedings of the 2007 workshop on Networked systems for developing regions*, pp. 1-6. 2007. https://doi.org/10.1145/1326571.1326585

[7] Eriksson, Jakob, Lewis Girod, Bret Hull, Ryan Newton, Samuel Madden, and Hari Balakrishnan. "The pothole patrol: using a mobile sensor network for road surface monitoring." In *Proceedings of the 6th international conference on Mobile systems, applications, and services*, pp. 29-39. 2008. https://doi.org/10.1145/1378600.1378605

[8] Rode, Sudarshan S., Shonil Vijay, Prakhar Goyal, Purushottam Kulkarni, and Kavi Arya. "Pothole detection and warning system: infrastructure support and system design." In *2009 International Conference on electronic computer technology*, pp. 286-290. IEEE, 2009. https://doi.org/10.1109/ICECT.2009.152

[9] Chang, K. T., J. R. Chang, and J. K. Liu. "Detection of pavement distresses using 3D laser scanning technology." In *Computing in civil engineering (2005)*, pp. 1-11. 2005. https://doi.org/10.1061/40794(179)103

[10] Li, Qingguang, Ming Yao, Xun Yao, and Bugao Xu. "A real-time 3D scanning system for pavement distortion inspection." *Measurement Science and Technology* 21, no. 1 (2009): 015702. https://doi.org/10.1088/0957-0233/21/1/015702

[11] Wang, Kelvin CP. "Challenges and feasibility for comprehensive automated survey of pavement conditions." In *Applications of advanced technologies in transportation engineering (2004)*, pp. 531-536. 2004. https://doi.org/10.1061/40730(144)99

[12] Hou, Zhiqiong, Kelvin CP Wang, and Weiguo Gong. "Experimentation of 3D pavement imaging through stereovision." In *International Conference on Transportation Engineering 2007*, pp. 376-381. 2007. https://doi.org/10.1061/40932(246)62

[13] Joubert, Deon, Ayanda Tyatyantsi, Jeffry Mphahlehle, and Vivian Manchidi. "Pothole tagging system." (2011).

[14] Moazzam, Imran, Khurram Kamal, Senthan Mathavan, S. Usman, and Mujib Rahman. "Metrology and visualization of potholes using the microsoft kinect sensor." In *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, pp. 1284-1291. IEEE, 2013. https://doi.org/10.1109/ITSC.2013.6728408

[15] Mahmoudzadeh, Ahmadreza, S. Firoozi Yeganeh, and Amir Golroo. "Kinect, a novel cutting edge tool in pavement data collection." *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 40 (2015): 425-431. https://doi.org/10.5194/isprsarchives-XL-1-W5-425-2015

[16] Lin, Jin, and Yayu Liu. "Potholes detection based on SVM in the pavement distress image." In *2010 Ninth International Symposium on Distributed Computing and Applications to Business, Engineering and Science*, pp. 544-547. IEEE, 2010. https://doi.org/10.1109/DCABES.2010.115

[17] Koch, Christian, and Ioannis Brilakis. "Pothole detection in asphalt pavement images." *Advanced engineering informatics* 25, no. 3 (2011): 507-515. https://doi.org/10.1016/j.aei.2011.01.002

[18] Buza, Emir, Samir Omanovic, and Alvin Huseinovic. "Pothole detection with image processing and spectral clustering." In *Proceedings of the 2nd International Conference on Information Technology and Computer Networks*, vol. 810, p. 4853. 2013.

[19] Ryu, Seung-Ki, Taehyeong Kim, and Young-Ro Kim. "Image-based pothole detection system for ITS service and road management system." *Mathematical Problems in Engineering* 2015 (2015): 1-10. https://doi.org/10.1155/2015/968361

[20] Huidrom, Lokeshwor, Lalit Kumar Das, and S. K. Sud. "Method for automated assessment of potholes, cracks and patches from road surface video clips." *Procedia-Social and Behavioral Sciences* 104 (2013): 312-321. https://doi.org/10.1016/j.sbspro.2013.11.124

[21] Koch, Christian, Gauri M. Jog, and Ioannis Brilakis. "Automated pothole distress assessment using asphalt pavement video data." *Journal of Computing in Civil Engineering* 27, no. 4 (2013): 370-378. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000232

[22] Tedeschi, Antonio, and Francesco Benedetto. "A real-time automatic pavement crack and pothole recognition system for mobile Android-based devices." *Advanced Engineering Informatics* 32 (2017): 11-25. https://doi.org/10.1016/j.aei.2016.12.004

[23] Bochkovskiy, Alexey, Chien-Yao Wang, and Hong-Yuan Mark Liao. "Yolov4: Optimal speed and accuracy of object detection." *arXiv preprint arXiv:2004.10934* (2020).

[24] Hui, Jonathan. “MAP (Mean Average Precision) for Object Detection.”Medium, Medium, 3 Apr. 2019.

[25] Redmon, Joseph, Santosh Divvala, Ross Girshick, and Ali Farhadi. "You only look once: Unified, real-time object detection." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779-788. 2016.

https://doi.org/10.1109/CVPR.2016.91
[26] Wang, Chien-Yao, Hong-Yuan Mark Liao, Yueh-Hua Wu, Ping-Yang Chen, Jun-Wei Hsieh, and I-Hau Yeh. "CSPNet: A new backbone that can enhance learning capability of CNN." In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pp. 390-391. 2020. https://doi.org/10.1109/CVPRW50498.2020.00203
[27] Wang, Tiancai, Rao Muhammad Anwer, Hisham Cholakkal, Fahad Shahbaz Khan, Yanwei Pang, and Ling Shao. "Learning rich features at high-speed for single-shot object detection." In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 1971-1980. 2019. https://doi.org/10.1109/ICCV.2019.00206
[28] Liu, Songtao, and Di Huang. "Receptive field block net for accurate and fast object detection." In *Proceedings of the European conference on computer vision (ECCV)*, pp. 385-400. 2018. https://doi.org/10.1007/978-3-030-01252-6_24