Article

Implementation of Pavement Defect Detection System on Edge Computing Platform

Yu-Chen Lin [1],*[ID]**, Wen-Hui Chen** [2] **and Cheng-Hsuan Kuo** [1]

[1] Department of Automatic Control Engineering, Feng Chia University, Taichung 40724, Taiwan; m0907354@o365.fcu.edu.tw
[2] Graduate Institute of Automation Technology, National Taipei University of Technology, Taipei 106344, Taiwan; whchen@ntut.edu.tw
* Correspondence: yuchlin@fcu.edu.tw

Abstract: Road surfaces in Taiwan, as well as other developed countries, often experience structural failures, such as patches, bumps, longitudinal and lateral cracking, and potholes, which cause discomfort and pose direct safety risks to motorists. To minimize damage to vehicles from pavement defects or provide the corresponding comfortable ride promotion strategy later, in this study, we developed a pavement defect detection system using a deep learning perception scheme for implementation on Xilinx Edge AI platforms. To increase the detection distance and accuracy of pavement defects, two cameras with different fields of view, at 70° and 30°, respectively, were used to capture the front views of a car, and then the YOLOv3 (you only look once, version 3) model was employed to recognize the pavement defects, such as potholes, cracks, manhole covers, patches, and bumps. In addition, to promote continuous pavement defect recognition rate, a tracking-via-detection strategy was employed, which first detects pavement defects in each frame and then associates them to different frames using the Kalman filter method. Thus, the average detection accuracy of the pothole category could reach 71%, and the miss rate was about 29%. To confirm the effectiveness of the proposed detection strategy, experiments were conducted on an established Taiwan pavement defect image dataset (TPDID), which is the first dataset for Taiwan pavement defects. Moreover, different AI methods were used to detect the pavement defects for quantitative comparative analysis. Finally, a field-programmable gate-array-based edge computing platform was used as an embedded system to implement the proposed YOLOv3-based pavement defect detection system; the execution speed reached 27.8 FPS while maintaining the accuracy of the original system model.

Keywords: pavement defect detection; YOLOv3; tracking-via-detection; Taiwan pavement defect image dataset (TPDID); edge computing; embedded system

1. Introduction

Road potholes and patches pose risks to road users; they frequently cause motorcycle riders to lose balance and fall. Generally, they are the major causes of damage to the chassis and suspension of vehicles. They can seriously endanger the lives of drivers and passersby. According to a report [1] by the Royal Automobile Club, a UK-based road rescue company, the quality of roads has deteriorated over the past decade, and the number of road rescue cases due to potholes has doubled. Road infrastructure is a significant factor in the productivity, safety, and satisfaction of a country. Research conducted by the European Community Research and Development Information Service [2] indicated that one-third of the traffic accidents in Europe are attributed to bad road conditions. A total of 20 million potholes were identified across Europe in 2011, but just half of them were patched, costing almost €1.2 billion. Even though automobile manufacturers have worked hard to reinforce chassis, it is still hard to withstand the ravages of the road conditions. For vehicles, potholes in the road may result in damage to the vehicle's shock absorber; the collision may trigger deformation and damage of the wheel rim. It may also cause the vehicle to be temporarily

out of control or even drift lanes due to a flat tire, causing a serious car accident. Moreover, the production boom of automated vehicles (AVs) is sweeping across the world. As of July 2019, there were 56 kinds of AVs in over 128 cities globally. Approximately 50% of these are shuttle buses for public transport. Besides safety, a crucial issue for the future of AVs is comfort, which is perhaps more important. Therefore, potholes on road pavements affect not only the structural safety of the vehicle but also the comfort of passengers in AVs. In severe cases, it can even cause vehicles to lose control and contribute to injuries and casualties.

Potholes, manhole covers, patches, and cracks are among the pavement defects identified in this study using an imaging lens and a deep learning (DL) YOLOv3 (you only look once, version 3) model [3]. In addition, we employed the Kalman tracking algorithm to improve the reliability and detection rate of the system, which not only significantly reduces the miss rate and increases the average precision (AP) but also stabilizes the DL prediction frame for surface defects. Moreover, to identify potholes at a longer distance, we used two lenses with different fields of view (FoVs): 30° and 70° lenses. The 30° FoV lens is mainly used for identifying potholes at a distance. However, when a pothole is close to the vehicle, a pothole object may be too large to be fully represented in the image due to a small viewing angle, resulting in misjudgment or lack of identification. Therefore, we also employed the 70° FoV lens, which can be used to compensate for pothole detection at near and medium distances. Using these two lenses simultaneously can reinforce the reliability of the recognition results. Finally, we conducted a quantitative analysis and comparison of DL object models with real-time object recognition, which includes YOLOv3 [3] used in this paper as well as YOLOv4 (YOLO, version 4) [4], MobileNet-YOLO [5], TF-YOLO (tiny fast YOLO) [6], and RetinaNet [7]. We ultimately chose YOLOv3 and implemented it with a field-programmable gate array (FPGA)-based hardware accelerator as an embedded system; the results show that YOLOv3 can still fulfill the criteria for real-time operation while maintaining the accuracy of the original system model.

The main contributions of the proposed pavement defect detection system are as follows:

- Two different FOV lenses and tracking-via-detection strategy are used to increase the detection accuracy rate, decrease the misjudgment rate, and lengthen the recognition distance efficiently.
- A lightweight YOLOv3 model is developed by the model compression technology and then implemented on an FPGA-based hardware accelerator edge computing platform, which not only satisfies the real-time computing requirements while maintaining the accuracy of the original system model but also becomes a competitive product in the automotive electronics market.
- To prove the effectiveness of the proposed pavement defect detection strategy, the first dataset for Taiwan pavement defects is established, which is called the Taiwan pavement defect image dataset (TPDID).
- In order to verify the robustness of the proposed road defect detection system, an extensive video recorded on the Pacific Northwest highway [8] is used. The experimental results show that the proposed system can effectively recognize potholes under severe conditions, such as water surface reflection interference or potholes with complex structures.

2. Related Works

The ability of object image recognition is the key technology at the core of this research. In this study, we installed two lenses with different FoVs on a vehicle to capture images of the road pavement ahead. We also used YOLOv3, a DL object recognition model, to detect road pavement defects in front of the vehicle. Road defects identified in this paper are mainly classified into six categories: road potholes, transverse cracks, longitudinal cracks, transverse patches, longitudinal patches, and manhole covers, where the pothole category is primarily defined as the pothole in which the pavement defects must surpass 10 cm and

the vehicle's wheels can fall into. Most related literature or commercial implementations employed high-resolution lidar, high-intensity infrared sensors, and other high-cost perception equipment for pothole detection [9–14]. Their main purpose is to provide road repair engineering units for subsequent repair by sensing whether the current road is defective. Keiji et al. [15] used a high-intensity near-infrared light-emitting diode mounted under the front bumper of a vehicle to measure whether the current road surface is dry, slippery, or icy to give the driver a hint about the current road conditions. However, only relatively close potholes could be identified, and this is prone to misinterpretation. Ground-penetrating radar (GPR) has been frequently used to detect potholes and cracks to understand a continuous profile of existing road conditions. However, GPR is very expensive and requires well-trained, experienced operators; moreover, the result shows the possibility of crack detection that depends on the layers of cracks [16]. In addition, various academic institutes and scholars have used picture recognition to identify road conditions as a consequence of the advancement in image processing and sensor technologies. However, the literature on the identification of road pavement defects is inadequate, especially in the first-person perspective for pavement defect detection. Lin et. al. [17] used a nonlinear support vector machine model with a Gaussian radial basis function to detect potholes. However, this scheme is limited by the ambient light source and weather factors, and it needs quite a bit closer recognizable distance. Fan et al. [18] used binocular vision for image capture of road pavements and identified the road potholes based on a parallax stereogram. This strategy, though, requires fairly close proximity to identify details regarding potholes. Moreover, when the illuminance is low at night or in adverse weather, the algorithm becomes invalid. Choudhury et al. [19] used a wide-angle lens mounted on the top of a vehicle to capture images of road pavement ahead. Simultaneously, a perspective projection transformation approach was used to detect road potholes in the aerial image using a blob recognition algorithm. This technique only provided improved detection performance for potholes at relatively short distances and is insufficient for all-weather applications. The watershed approach detection of road potholes was also used by Chung et al. [20] in detecting road potholes with smooth, broken, and cracked pavements. However, the recognition rate is not high and is limited by the ambient light source. Recently, DL techniques have been achieving state-of-the-art results for object detection on standard benchmark datasets and computer vision competitions. In 2019, Chun et al. [21] employed fully convolutional neural networks with semantic segmentation efficacy and virtual samples generated by autoencoders. They used a semisupervised learning approach to conduct road pavement hazard detection. Moreover, Baek et al. [22] proposed a pothole classification model using edge detection and a YOLO-based feature extraction scheme for pothole detection. However, these studies have only considered ideal situations such as daytime on a sunny day. The weather variations or illumination changes, such as poor weather conditions, shadows cast from objects or water, and low-illumination environment, are challenging problems. In addition, there are only subjective consciousness assessments lacking objective quantitative analysis. Consequently, no effective approach is presently accessible for detecting road pavement defects, and most significantly, pavement defects that are further out cannot be detected so as to provide further application in vehicle control, such as a pothole warning system or vehicle stability control system, and to simultaneously meet all-weather requirements. Besides, for vehicle applications, the overly complicated DL network architecture is unable to meet the requirements of a real-time system and embedded hardware with vehicle specifications.

Although current object recognition algorithms based on deep neural networks have outstanding performance in many applications, DL models often need high-performance computing hardware. e.g., multicore high-performance graphics processing unit and a similar processing system, and consume a lot of power; thus, several AI-based applications are difficult to apply to embedded systems with limited memory space and computing resources. Therefore, in this study, in addition to the development of an AI-based defect detection scheme, we use the model compression technology toward a lightweight model,

and then an FPGA-based hardware accelerator edge computing platform is employed as an embedded system to implement the proposed pavement defect detection model. Experimental results show that the proposed system not only satisfies the real-time computing requirements while maintaining the accuracy of the original system model but also becomes a competitive product in the automotive electronics market.

3. Pavement Defect Detection Scheme

In this study, YOLOv3 was employed as the primary DL model to detect road defects. Two lenses with different FoVs, 30° and 70°, were used for anterior image acquisition. By using the two FoV cameras, minor pavement defects, including tiny potholes or patches, may also be detected completely and accurately at a long distance. Moreover, pavement defects at close proximity are detected based on the results of a large-angle lens, which will not lead to misjudgment, as the small-angle lens is unable to capture the full image of a pavement defect when it is too close. Particularly, the 30° FoV lens is used for identifying long-distance pavement defects, whereas the 70° FoV lens is used primarily to locate medium and proximal defects in road pavement.

Consequently, in the proposed model, the application of both recognition results not only effectively increases the detection accuracy rate but also reduces misjudgments and improves reliability. The structure of the proposed road pavement defect detection model is shown in Figure 1. In addition, a Kalman filter (KF) was integrated to improve system reliability and robustness to ensure continuity and reliability of the detection model. The outcome of the quantitative analysis is presented in the experimental results in Section 5.
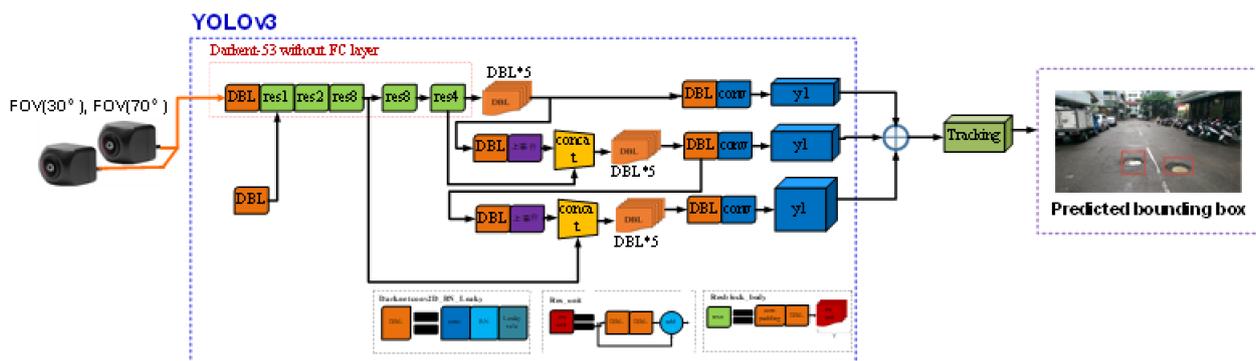

Figure 1. YOLOv3-based model for pavement defect detection.

Further, as the vehicle moves forward, the potholes get closer to the vehicle, and the pavement defect will become larger, which may induce misjudgment, e.g., marking several specific defects within the pavement defect or marking many tiny potholes within the pothole. In this scenario, we use a tracking-via-detection framework [23,24] (Figure 2). The KF [11,25] tracking algorithm with intersection over union (IOU) estimation is used to maintain the reliability of the detection model and to lower the miss rate. It primarily tracks the location, length, and width information of the object's prediction frame to improve the recognition rate.

The Kalman filter is used to update the estimate of the state variables by using the previous state estimate and the current state observation to obtain the current state estimate. During the approximation process, it is necessary to meet the minimum mean-squared error estimate so that the estimated value will be closer to the actual system state after each iteration. Thus, firstly, the state of the previous time step is used to predict the state of the next time step, which is the state prediction stage. Assume the variation of the prediction frame is linear, so the system state and output equation can be defined as

$$\hat{x}_{k|k-1} = A\hat{x}_{k-1|k-1} + Bu_{k-1},$$
$$z_k = H\hat{x}_{k|k-1} \tag{1}$$

where $\hat{x}_{k|k-1}$ is used to predict the estimate at time step k from the previous time step $(k-1)$; $\hat{x}_{k-1|k-1}$ is the best-estimated state at the previous time step $(k-1)$; u_{k-1} is the state control quantity at the previous time step $(k-1)$, which is assumed to be 0. Moreover, $z_k = \begin{bmatrix} x & y & h & w \end{bmatrix}^T$ is the output state of the system measurement at time step k, and it also refers to the result of the current pothole recognition prediction frame. Among them, x and y are the upper-left coordinates of the prediction frame, respectively, h and w are the length and width of the prediction frame; A is the state transition matrix; B is the control matrix. A is extremely important, characterized in particular by the dynamic relationship between the pavement defect prediction frame identified and the vehicle itself. The prediction of the prediction frame is currently assumed to be a linear relationship between speed and time. Subsequently, the covariance matrix of the estimated value is to be updated, which can be expressed as

$$P_{k|k-1} = AP_{k-1|k-1}A^T + Q \tag{2}$$

where $P_{k|k-1}$ is the covariance matrix at time step k predicted by the covariance matrix at time step $k-1$; Q is the system noise. The optimum state approximation outcome can be reached, and it can be represented in conjunction with the system state and observed values in [11] (i.e., the prediction frame details of the current pothole recognition result).

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k(z_k - H\hat{x}_{k|k-1}) \tag{3}$$

where K_k is the Kalman gain, which represents the weighting between the measured value and the estimated value. It also indicates that when the value of K_k is large, it means the credit of observed value is greater than the estimated value currently; otherwise, it means that the credit of estimated value is greater than the observed value. Its Kalman gain can be estimated by the following equation:

$$K_k = P_{k|k-1}H^T(HP_{k|k-1}H^T + R)^{-1} \tag{4}$$

where R is the observation error matrix, and the covariance matrix at the last time step k can be updated to

$$P_{k|k} = (I - K_kH)P_{k|k-1} \tag{5}$$

We can estimate the current state (also referred to as the coordinate, length, and width details of the pavement defect prediction frame) based on the state of the prior time step and modify the predictive information with the observation information to obtain the best estimate. Therefore, this paper seizes on the Kalman filter, which may not only prevent the sudden emergence of no prediction frame due to misjudgment but also be less vulnerable to error information due to sudden misjudgment in the prediction frame of each identification outcome. The detection block in Figure 2 is the identification of defects in road pavement based on the YOLOv3 model used in this study. In order to increase the computational efficiency, the image of 30° FOV lens and the image of 70° FOV lens were merged into one image via coordinate transformation, and this merged image was used as the input to the YOLOv3 model to identify pavement defects. At this point, it is the identification result when $t = 1$. Then, we used the Kalman filter to track the identified road pavement defect prediction frame for $t = 2 \sim j$. However, we must simultaneously mark the object at time step $t = j$, to prevent any tracking loss that may happen as a consequence of the object's absence or masking, and compare the identification outcome to the tracking consequence by IOU. When IOU > 0.5, the tracking results can continue to be trustworthy; otherwise, the new identification results can be used as the basis of subsequent tracking. This process will increase the recognition rate efficiently and reduce the miss rate.
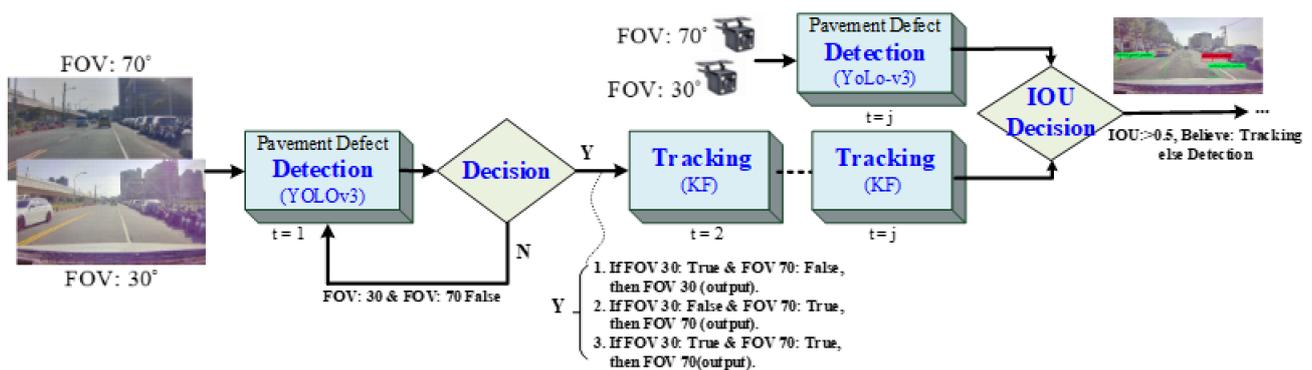
Figure 2. Tracking-via-detection framework.

4. Lightweight Model and Embedded System Implementation

At present, most recognition systems based on DL technology are often computationally complex; thus, they need to be built on the basis of high-performance computing equipment. However, these high-performance computing platforms not only are expensive but also have the disadvantage of high power consumption and heat dissipation problems. Thus, these platforms cannot be used in the automotive industry. To overcome the above problems, the model compression technology was used in the proposed pavement defect detection model, and then the lightweight model was deployed on the Xilinx ZCU104 embedded system.

The purpose of model compression is to achieve a lightweight model that is simplified from the original without significantly reducing accuracy. Various model compression strategies have been proposed, including microarchitecture [26,27], pruning [28–30], quantization [31,32], low-rank decomposition [33,34], and knowledge distillation [35,36]. Each scheme has its advantages and disadvantages as well as suitable situations. For example, the microarchitecture method mainly simplifies the model structure to reduce the number of network parameters; however, it usually sacrifices a certain level of accuracy. Although some unstructured pruning or sparsifying methods can accurately remove insignificant parameters to achieve model compression, they cannot be accelerated by general hardware. In this study, quantization was used for the YOLOv3 model; the flowchart of model compression and verification is shown in Figure 3. The model weight of the original YOLOv3 training framework was a 32-bit floating-point representation with 246.3-MB size. Through the quantification process, the model weight was converted to an 8-bit integer representation. The quantified model was only 23% of the original model's size. The improvement effect of the model compression ratio is shown in Figure 4. For the quantified model, part of training data were used for fine-tuning again to restore the detection accuracy of the original model. Therefore, the lightweight model did not lose its accuracy when the model compression ratio became 77%.
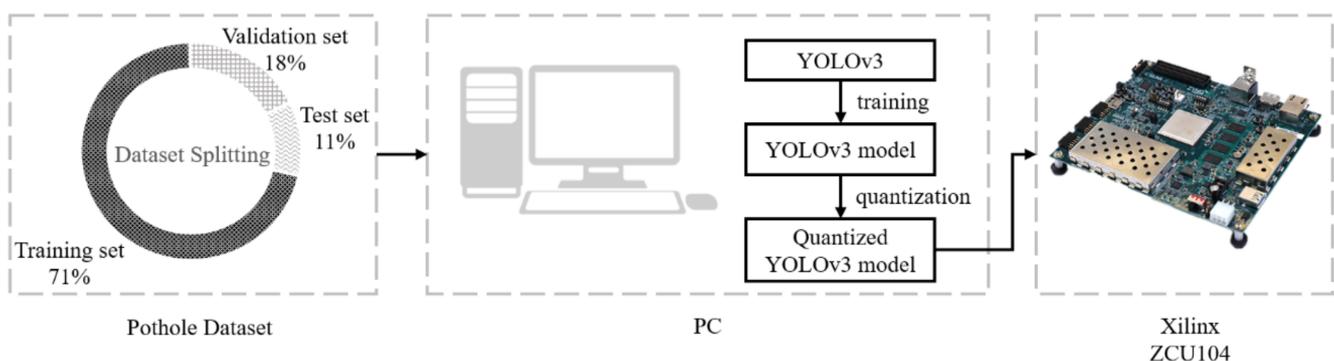

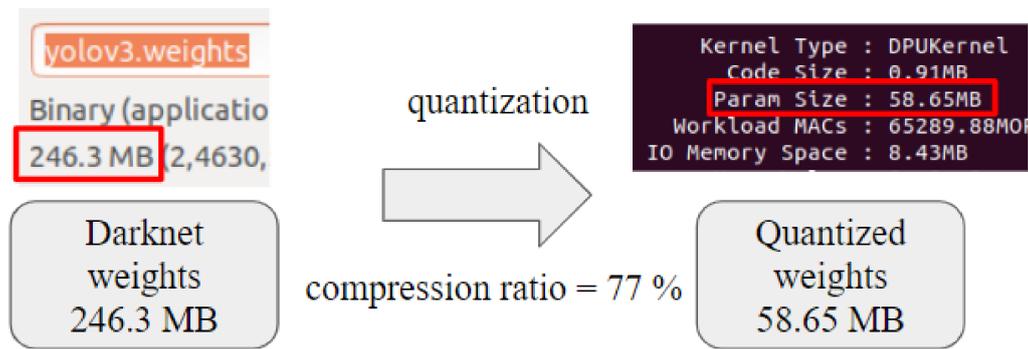**Figure 3.** Model compression strategy and verification flowchart.

Figure 4. Improvement effect of model compression ratio.

Thus, the quantified model could be implemented on the Xilinx FPGA-based embedded platform for the proposed real-time pavement defect detection model to decrease the requirements of computational resources and satisfy the low power consumption requirement of the automotive electronics industry. As shown in Figure 5, two AI-based embedded systems were used for comparing the quantitative performance: one was Nvidia TX2 using graphics processing unit acceleration scheme, and the other was Xilinx ZCU104 platform via FPGA hardware acceleration. Figure 5c shows the quantitative performance comparison for the two embedded systems, when the input image size was 416×416 and the average accuracy of the PC-based identification system before compression was maintained. The execution speed of the Xilinx ZCU104 embedded system was 27.8 FPS, which not only achieved the real-time demand response but also satisfied the low power consumption requirement of the automotive electronics industry.
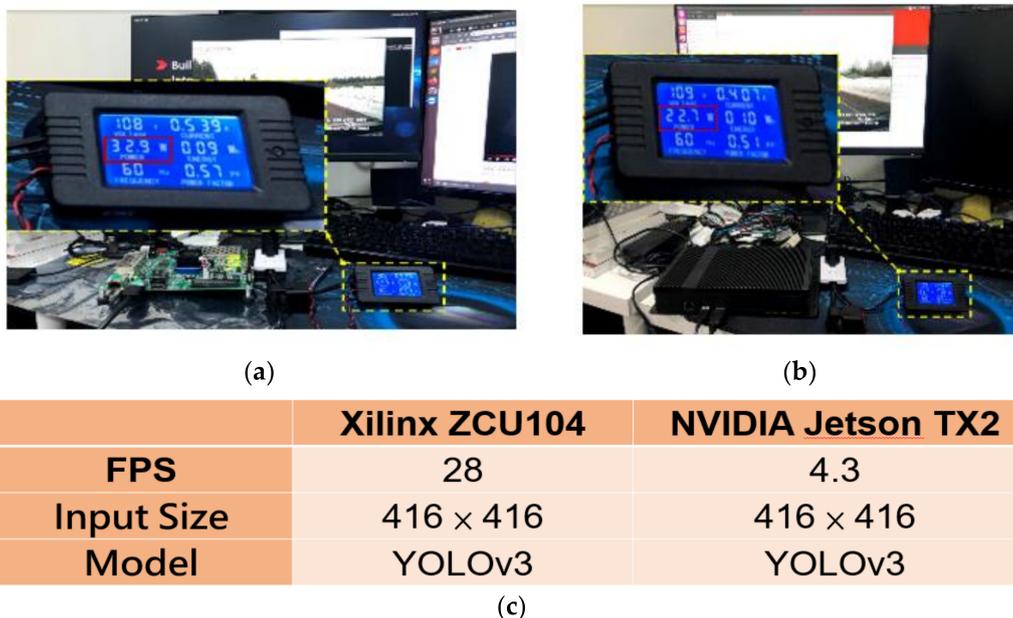

(**a**)

(**b**)

	Xilinx ZCU104	NVIDIA Jetson TX2
FPS	28	4.3
Input Size	416×416	416×416
Model	YOLOv3	YOLOv3

(**c**)

Figure 5. Implementation of embedded system for YOLOv3-based pavement defect detection systems. (**a**) Xilinx ZCU104 platform; (**b**) Nvidia TX2 platform; (**c**) quantitative performance comparison.

5. Experimental Results and Quantitative Analysis

To prove the effectiveness of the proposed detection strategy, experiments were performed on our established Taiwan pavement defect image dataset (TPDID), which is the first dataset for Taiwan pavement defects (Figure 6). TPDID has 62,000 samples, and pavement defects are divided into six categories, namely Category 0: road potholes and sudden slopes (31,708 samples), Category 1: manhole covers (9284 samples), Category

2: longitudinal cracks (16,810 samples), Category 3: transverse cracks (4761 samples), Category 4: longitudinal patches (18,023 samples), and Category 5: transverse patches (10,291 samples). The lenses used for this database were RGB three-channel lenses that complied with vehicle application regulations. In this study, two lenses with different FoVs were used to achieve efficient recognition at short, medium, and long distances; this strategy can efficiently increase the detection accuracy rate, decrease the misjudgment rate, and lengthen the recognition distance.
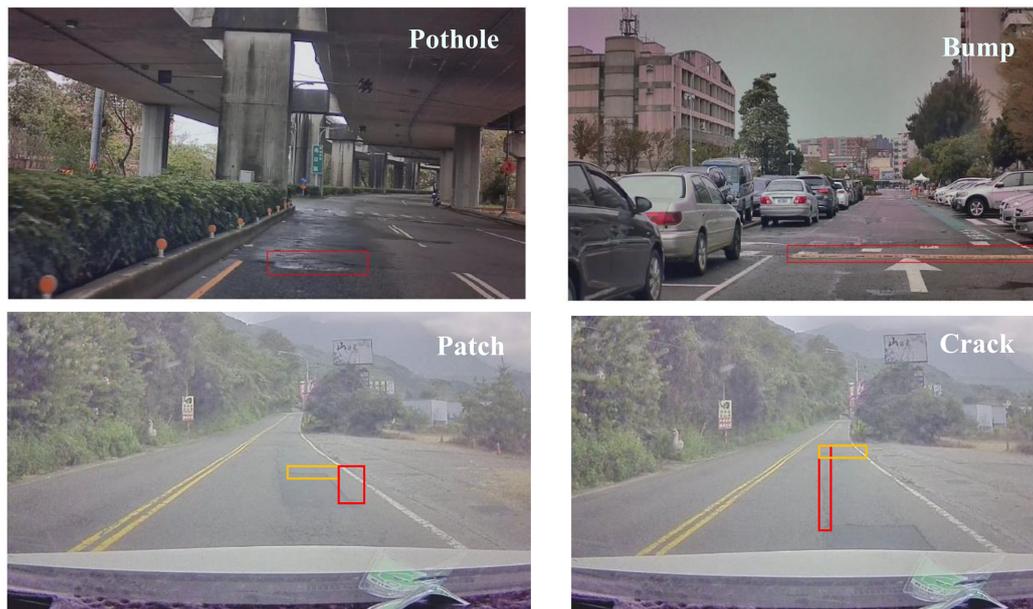
Figure 6. Taiwan pavement defect image database (TPDID).

Figure 7 shows the results of pavement defect detection using the YOLOv3 model. The images of the road pavement captured via the 30° FoV and 70° FoV lenses are on the left and right sides of Figure 7, respectively. The top two images in Figure 7a,b show the detection results of road pavement defects at time step *t*. The image below shows the detection and tracking results at time step *t + j*. As the left side of Figure 7 indicates that 30° FoV has a focal length in the distance due to the small angle of view, the pavement defects in the far FoV are large and clear. The right side of 70° FoV has a broader viewing angle, wider viewing field, and closer focal length, making it suitable for detecting medium and short distance objects. The pink frame in Figure 7 is the identified pothole category, the red frame is the manhole cover category, and the green frame is the longitudinal crack. Alongside this, we can observe from the upper left of Figure 7a that when the road pavement defect is far away from the vehicle, it can be identified by 30° FoV. By contrast, no detection results are available via 70° FoV as shown in the upper right of Figure 7a. With this in mind, the detection results of 30° FoV are preferable when smaller pavement defects are found; moreover, if the distance between the pavement defects and the vehicle is decreasing, the detection results of 70° FoV are preferred to avoid incomplete pavement defects. In addition, since the features of the road potholes are not noticeable in Figure 7b, no recognition results are obtained from the images of the two distinct FoV perspectives at time step *t* (no pink prediction frame at the top of Figure 7b). Conversely, at time step *t + j*, when the distance between the pothole and the vehicle is decreasing (the two lower images in Figure 7b), both images identify the pothole. Thus, the approach presented in this study not only effectively extends the recognition distance of pavement defects using two different FoV lenses but also contributes to effectively increase the detection accuracy rate and decrease the miss rate. Meanwhile, a tracking algorithm was employed in this study to improve the reliability of the detection system.

Figure 7. Pavement defect detection results by two lenses with different viewing angles. (**a**) Detection results at time step t; (**b**) Detection and tracking results at time step *t + j*.

Figure 8 shows the loss function and mean AP (mAP) of the neural network training process of YOLOv3 for the proposed pavement defect detection system. As seen in Figure 8, we consider the network model parameter set at the maximum mAP (approximately 70.2415%) as the network model applied for testing. Figure 9 demonstrates the quantitative analysis for six pavement defect categories using the YOLOv3 model. The left-hand side of Figure 9a indicates the average detection accuracy of each category, and the right-hand side of Figure 9b shows the miss rate of each category. Notably, when the IOU threshold value >0.5, its pothole category average precision is about 71%, and the miss rate is 29%, according to the quantification result. The best detection results were obtained from Category 1, which had an average precision of 89% and a miss rate of 17%. This is primarily because there is a considerable amount of manhole covers on the roads of Taiwan, and the features are very distinctive, which is why the detection performance of this category is the best.
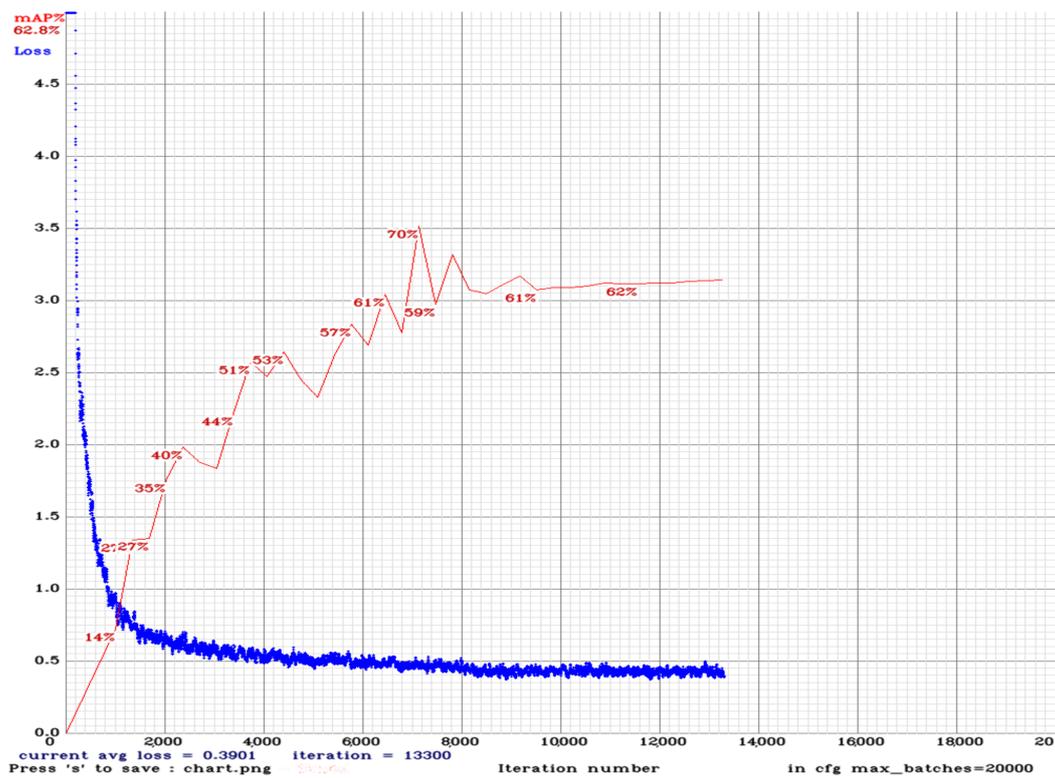
Figure 8. Loss function and mAP for our proposed pavement defect detection system.
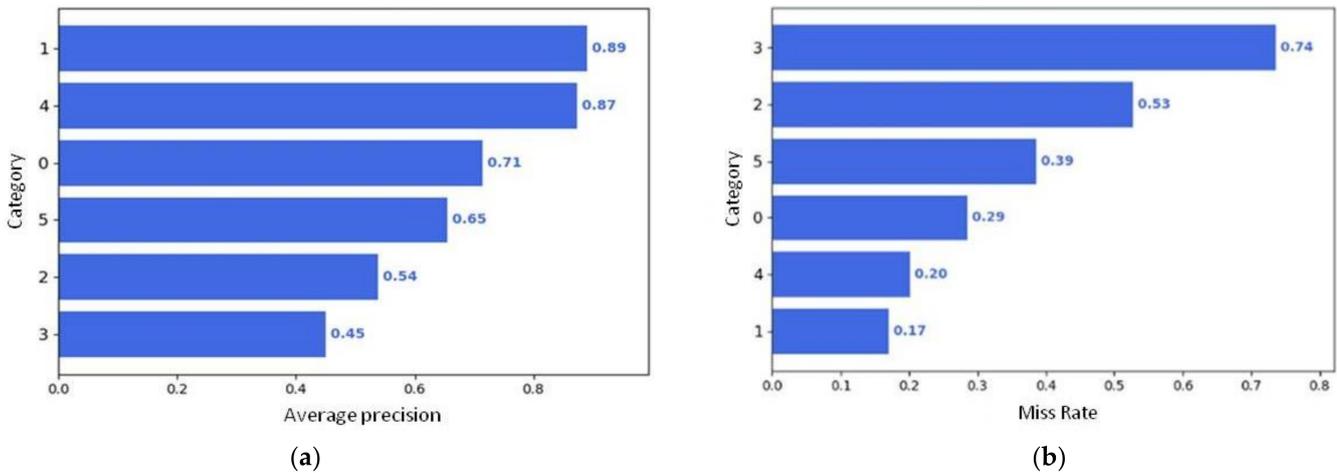

(**a**)

(**b**)

Figure 9. Average precision and miss rate for 6 pavement defect categories. (**a**) Average detection accuracy of each category; (**b**) Miss rate of each category.

Figure 10 shows the pavement defect detection results for Category 0 under a low-illumination night environment via a car camera. Since an average car camera has a high dynamic range characteristic, it can be seen from Figure 10 that the output image looks like a grayscale image. Although the image is no color information, it can effectively reduce the noise for the camera in low-light conditions and further improve the ability to identify objects at night. As shown in Figure 10, it can still give a great pothole recognition effect in a low-light environment at night.

Figure 10. Pothole recognition results in low-illumination night environment.

To verify the robustness of the proposed road defect detection system, an extensive video recorded on the Pacific Northwest highway [36], which includes the potholes and patches of snow and water, was used. The experimental results show that the proposed system can effectively recognize potholes under severe conditions, such as water surface reflection interference or potholes with complex structures (see Figure 11 below). The experimental results can be visualized via videos that can be found at the following link: https://youtu.be/_6OhNANpTWg (accessed on 4 February 2021).
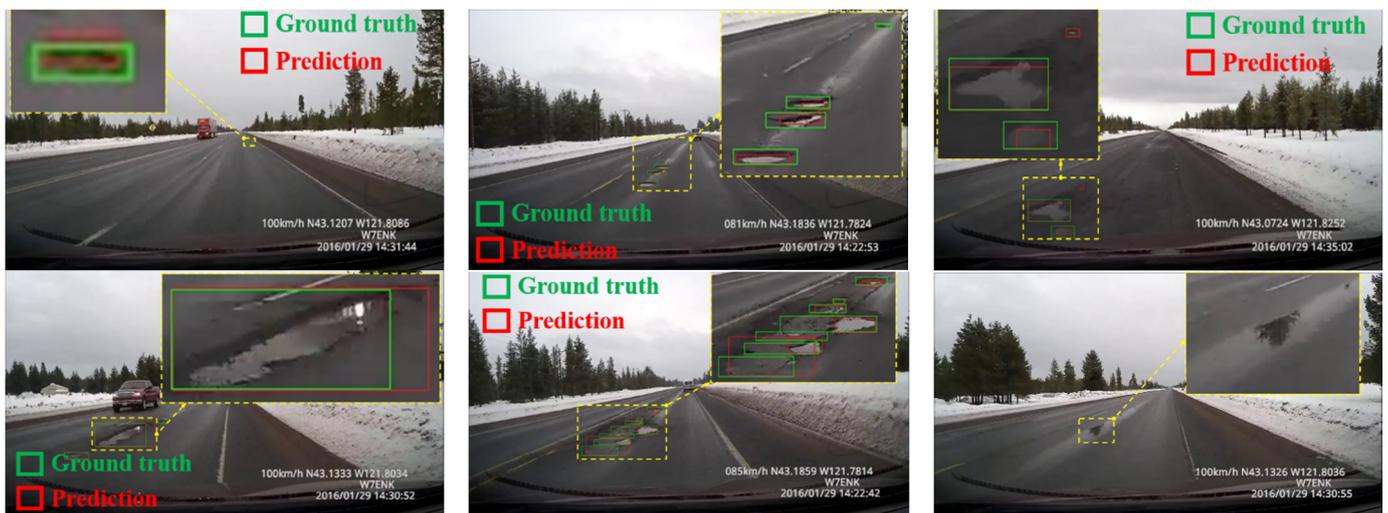

Figure 11. Pothole recognition results under severe conditions from the Pacific Northwest highway dataset.

We also conducted the quantitative analysis and comparison of prominent PC-based DL object recognition models, such as YOLOv4, MobileNet-YOLO, TF-YOLO, and RetinaNet, for pavement defect detection. Figure 12 shows the detection results for different models for pavement defects, where panels a–d show the detection results of YOLOv4, MobileNet-YOLO, TF-YOLO, and RetinaNet, respectively. YOLOv4 had good detection results, whereas MobileNet-YOLO and TF-YOLO had poor detection results for 30° FoV, especially TF- YOLO, and the recognition qualities were not good with both FoVs. Lastly, RetinaNet could also identify some defects, but some defects were not identified.

Figure 12. Pavement defect detection results using different deep learning models. (**a**) YOLOv4; (**b**) YOLO-mobilenet; (**c**) TF-YOLO; (**d**) RetinaNet.

Further, we conducted a quantitative comparison and analysis of mAP and FPS for each of the abovementioned models (Figure 13). From the figure, except for TF-YOLO, whose computational speed attained roughly 27 FPS, all other models reached over 30 FPS, which fulfilled the demand of real-time vehicle applications. MobileNet-YOLO had the fastest computation speed, but the average detection accuracy rate was relatively low; meanwhile, YOLOv4 not only had a higher detection score but also fulfilled the demands of real-time operation. In consideration of both mAP and computing speed, we adopted the YOLOv3 model with the KF tracking algorithm as the road pavement defect detection algorithm and implemented the system via Xilinx embedded hardware platform. We will attempt to use the YOLOv4 model with another tracking algorithm in the future and believe that it will have a much better recognition effect.
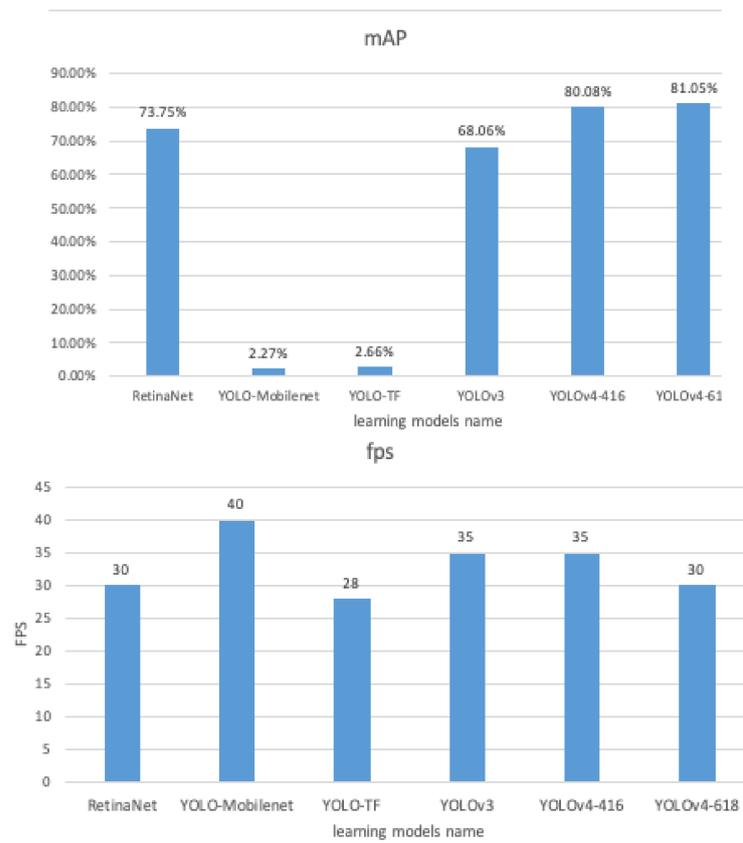
Figure 13. Quantitative comparison of mAP and FPS for different deep learning models.

Figure 14 demonstrates the miss rate analysis of the DL object detection model compared in this study using TPDID. It is obvious that YOLOv4 focused on the condition of 10^{-1} FPPI and had a smaller miss rate, implying that the model corresponds effectively with the proposed road pavement defect detection system. In the future, we will continue to build up sample dataset repositories to further enhance the detection accuracy and reliability of the proposed detection system.
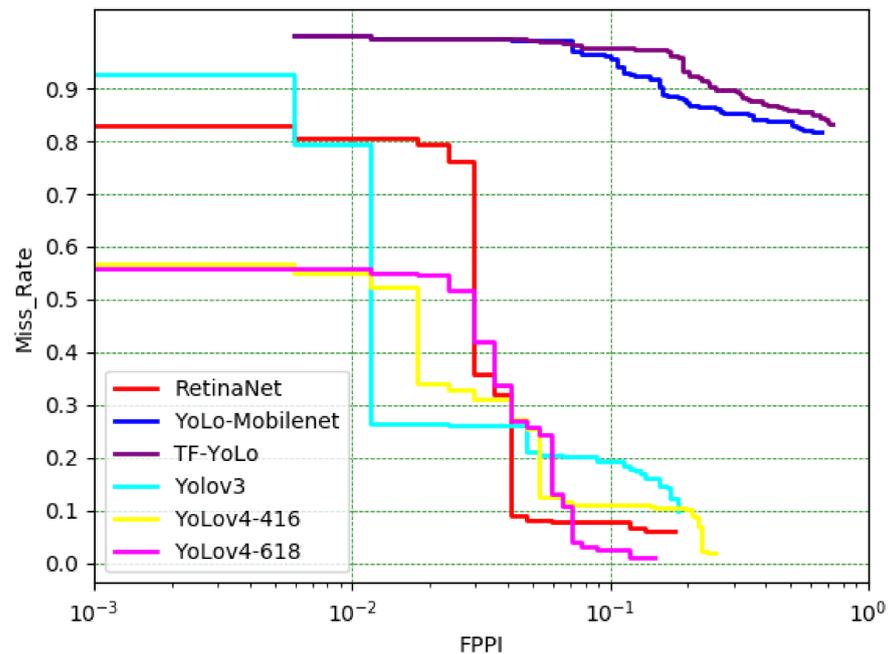

Figure 14. Comparison of miss rate for different deep learning models.

Finally, the AP of Category 0 for each model in different distance intervals with respect to 70° FoV was evaluated. The results are shown in Figure 15; when the defect is approximately 20–30 m away from the vehicle, the vertical (*Y*-axis) size of the image is about 400 pixels. YOLO series resulted in a better detection accuracy rate than others when IOU is 0.5, especially in the interval of 0–2, i.e., within 0–16 m, which means that the YOLO series can have better detection results at both medium and short distances.
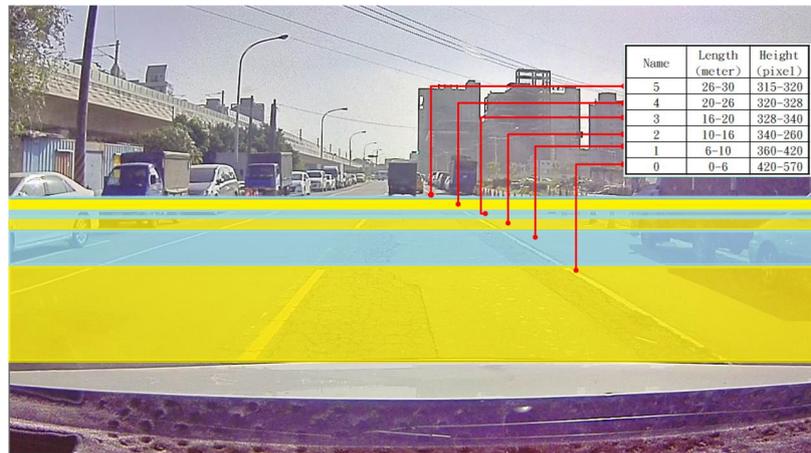
(**a**)
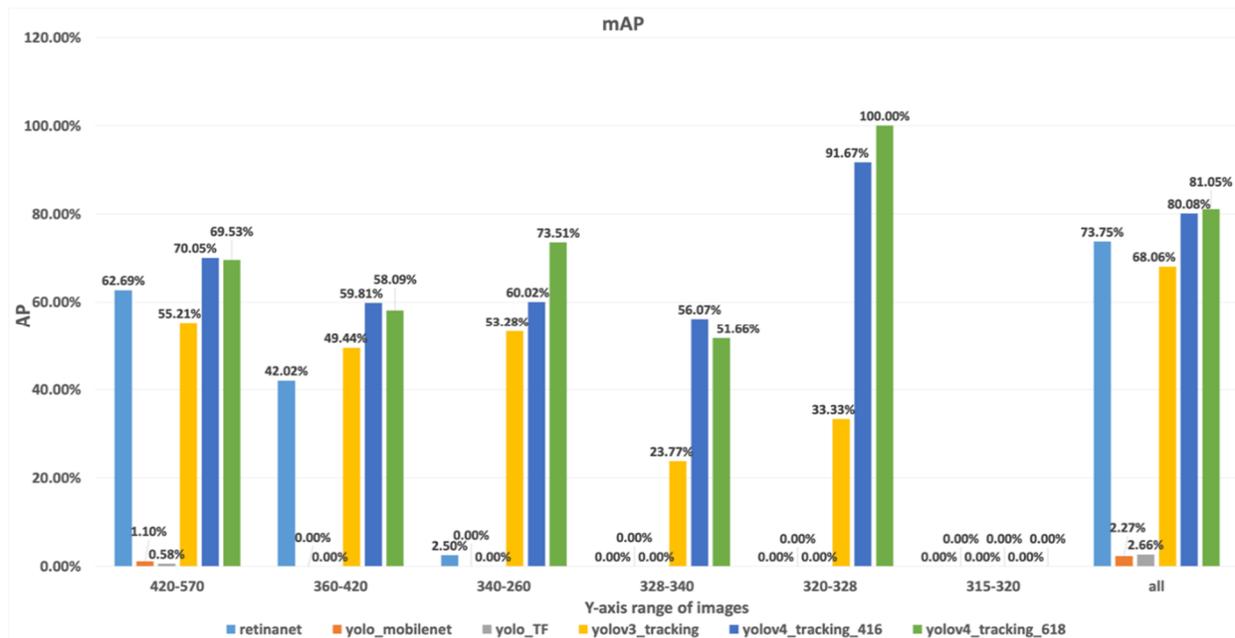
(**b**)

Figure 15. Quantitative comparison of AP of pothole category for each model in different distance intervals. (**a**) Relationship between the pixel differences to actual road distance; (**b**) AP evaluation results for each model in different distance intervals.

6. Conclusions

In this study, to detect road pavement defects in front of vehicles, two cameras with different FoVs were employed. Further, YOLOv3 and KF were integrated to effectively enhance the detection accuracy and lower the miss rate. We determined that the detection accuracy for the road pothole category reached 71%, with a 29% miss rate. In addition, we conducted a lightweight design of the YOLOv3 model via network weight quantization technology and successfully implemented it on commercially available embedded hardware with FPGA acceleration functionality. It can run at a speed of 27.8 FPS to fulfill

the requirement of automotive applications. Moreover, we also constructed an image repository of the Taiwan pavement defect image database (TPDID) and, finally, quantitatively analyzed and compared the modern real-time DL object recognition models. In the future, we will continue to build up the sample data under different weather and ambient brightness conditions to satisfy the robustness requirements for all weather conditions.

References

1. Royal Automobile Club (RAC). *Report a Pothole and Find out How to Claim for Damage*; RAC Foundation: London UK, 2018.
2. World Economic Forum. *Ranking of the Countries with the Highest Road Quality in 2017/2018*; Statista: Hamburg Germany, 2018.
3. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767.
4. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934.
5. Li, Y.; Han, Z.; Xu, H.; Liu, L.; Li, X.; Zhang, K. YOLOv3-Lite: A Lightweight Crack Detection Network for Aircraft Structure Based on Depthwise Separable Convolutions. *Appl. Sci.* **2019**, *9*, 3781. [CrossRef]
6. He, W.; Huang, Z.; Wei, Z.; Li, C.; Guo, B. TF-YOLO: An Improved Incremental Network for Real-Time Object Detection. *Appl. Sci.* **2019**, *9*, 3225. [CrossRef]
7. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollar, P. Focal Loss for Dense Object Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 318–327. [CrossRef] [PubMed]
8. Pacific Northwest Highway Dataset. Available online: www.youtube.com/watch?v=BQo87tGRM74 (accessed on 15 January 2020).
9. Bhatia, Y.; Rai, R.; Gupta, V.; Aggarwal, N.; Akula, A. Convolutional Neural Networks Based Potholes Detection Using Thermal Imaging. *J. King Saud Univ. Comput. Inf. Sci.* **2019**, 1–11, In press.
10. Zeng, Q.; Wen, G.; Li, D. Multi-Target Tracking by Detection. In Proceedings of the 2016 International Conference on Audio, Language and Image Processing (ICALIP), Shanghai, China, 11–12 July 2016; pp. 370–374.
11. Kaur, H.; Sahambi, J.S. Vehicle Tracking in Video Using Fractional Feedback Kalman Filter. *IEEE Trans. Comput. Imaging* **2016**, *2*, 550–561. [CrossRef]
12. Kumar, A.; Kalita, D.J.; Singh, V.P. A Modern Pothole Detection Technique Using Deep Learning. In Proceedings of the 2nd International Conference on Data, Engineering and Applications (IDEA), Bhopal, India, 28–29 February 2020; pp. 1–5.
13. Srinidhi, G.; S-M, D.R. Pothole Detection Using CNN and AlexNet. In Proceedings of the International Conference on Communication and Information Processing (ICCIP-2020), Tokyo, Japan, 27–29 November 2020; pp. 1–9.
14. Ravi, R.; Habib, A.; Bullock, D. Pothole Mapping and Patching Quantity Estimates Using LiDAR-Based Mobile Mapping Systems. *Transp. Res. Rec.* **2020**, *2674*, 124–134. [CrossRef]
15. Fujimura, K.; Sakamoto, T. Road Surface Sensor. *Fujitsu TEN Technol. J.* **1988**, *1*, 64–72.
16. Fernandes, F.M.; Pais, J.C. Laboratory Observation of Cracks in Road Pavements with GPR. *Constr. Build. Mater.* **2017**, *154*, 1130–1138. [CrossRef]
17. Lin, J.; Liu, Y. Potholes Detection Based on SVM in the Pavement Distress Image. In Proceedings of the 2010 Ninth International Symposium on Distributed Computing and Applications to Business, Engineering and Science, Hong Kong, China, 10–12 August 2010; pp. 544–547.
18. Fan, R.; Ozgunalp, U.; Hosking, B.; Liu, M.; Pitas, I. Pothole Detection Based on Disparity Transformation and Road Surface Modeling. *IEEE Trans. Image Process.* **2019**, *29*, 897–908. [CrossRef]
19. Choudhury, A.; Ramchandani, R.; Shamoon, M.; Khare, A.; Kaushik, K. An Efficient Algorithm for Detecting and Measure the Properties of Pothole. In Proceedings of the Emerging Technology in Modelling and Graphics, Kolkata, India, 6–8 September 2018; pp. 447–457.
20. Chung, T.D.; Khan, M.K.A.A. Watershed-Based Real-Time Image Processing for Multi-Potholes Detection on Asphalt Road. In Proceedings of the 2019 IEEE 9th International Conference on System Engineering and Technology (ICSET), Shah Alam, Malaysia, 7 October 2019; pp. 268–272.

21. Chun, C.; Ryu, S.-K. Road Surface Damage Detection Using Fully Convolutional Neural Networks and Semi-Supervised Learning. *Sensors* **2019**, *19*, 5501. [CrossRef] [PubMed]
22. Baek, J.-W.; Chung, K. Pothole Classification Model Using Edge Detection in Road Image. *Appl. Sci.* **2020**, *10*, 6662. [CrossRef]
23. Chari, V.; Lacoste-Julien, S.; Laptev, I.; Sivic, J. On Pairwise Costs for Network Flow Multi-Object Tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 12 June 2015; pp. 5537–5545.
24. Schulter, S.; Vernaza, P.; Choi, W.; Chandraker, M. Deep Network Flow for Multi-Object Tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 26 July 2017; pp. 6951–6960.
25. Li, X.; Wang, K.; Wang, W.; Li, Y. A Multiple Object Tracking Method Using Kalman Filter. In Proceedings of the 2010 IEEE international conference on information and automation, Harbin, China, 20–23 June 2010; pp. 1862–1866.
26. Han, S.; Mao, H.; Dally, W.J. Deep Compression: Compressing Deep Neural Networks with Pruning, Trained Quantization and Huffman Coding. In Proceedings of the International Conference on Learning Representations, San Juan, Puerto Rico, 2–4 May 2016; pp. 1–14.
27. Hu, P.; Peng, X.; Zhu, H.; Aly, M.M.S.; Lin, J. OPQ: Compressing Deep Neural Networks with One-Shot Pruning-Quantization. In Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI-21), Vancouver, VN, Canada, 2–9 February 2021; pp. 1–9.
28. Zhang, T.; Ye, S.; Zhang, K.; Tang, J.; Wen, W.; Fardad, M.; Wang, Y. A Systematic DNN Weight Pruning Framework Using Alternating Direction Method of Multipliers. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 191–207.
29. Sun, X.; Ren, X.; Ma, S.; Wang, H. MeProp: Sparsified Back Propagation for Accelerated Deep Learning with Reduced Overfitting. In Proceedings of the 34th International Conference on Machine Learning (ICML), Sydney, Australia, 6–11 August 2017; pp. 1–10.
30. He, Y.; Ding, Y.; Liu, P.; Zhu, L.; Zhang, H.; Yang, Y. Learning Filter Pruning Criteria for Deep Convolutional Neural Networks Acceleration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 2006–2015.
31. Zhou, A.; Yao, A.; Guo, Y.; Xu, L.; Chen, Y. Incremental Network Quantization: Towards Lossless CNNs with Low-Precision Weights. In Proceedings of the International Conference on Learning Representations (ICLR), Toulon, France, 24–26 April 2017; pp. 1–14.
32. Wiedemann, S.; Kirchhoffer, H.; Matlage, S.; Haase, P.; Marban, A.; Marinč, T.; Neumann, D.; Nguyen, T.; Schwarz, H.; Wiegand, T. Deepcabac: A Universal Compression Algorithm for Deep Neural Networks. *IEEE J. Sel. Top. Signal Process.* **2020**, *14*, 700–714. [CrossRef]
33. Lin, S.; Ji, R.; Chen, C.; Tao, D.; Luo, J. Holistic CNN Compression via Low-Rank Decomposition with Knowledge Transfer. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *41*, 2889–2905. [CrossRef] [PubMed]
34. Lin, M.; Ji, R.; Wang, Y.; Zhang, Y.; Zhang, B.; Tian, Y.; Shao, L. Hrank: Filter Pruning Using High-Rank Feature Map. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 1529–1538.
35. Yim, J.; Joo, D.; Bae, J.; Kim, J. A Gift from Knowledge Distillation: Fast Optimization, Network Minimization and Transfer Learning. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 7130–7138.
36. Yin, H.; Molchanov, P.; Alvarez, J.M.; Li, Z.; Mallya, A.; Hoiem, D.; Jha, N.K.; Kautz, J. Dreaming to Distill: Data-Free Knowledge Transfer via Deepinversion. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 8715–8724.