Journal of Computer Science and Artificial Intelligence
ISSN: 3078-8242 (Print), ISSN: 3078-2309 (Online) | Vol. 6, No. 3, 2026

Application and Implementation of Deep Learning-Based Road Pothole Detection and Segmentation Algorithms

Shicen Liu, Junyong Liang [*], Meixue Lai, Yingmei Liao, Hao Tang, Bo Zhang

School of Civil Engineering, Sichuan University of Science and Engineering, Zigong 643000, China

[*] **Corresponding author**: (Email: 906401565@qq.com)

Abstract: Road pit detection is one of the important means to maintain road and traffic safety. Road pit detection includes target detection of pit holes in the video based on the driving recorder, and judging the location and size of road pit holes. This study aims to accurately judge the pit holes on the road, the size of the pits and the pit levels of the pits based on the driving recorder. Provide valuable pit repair advice for road inspectors, thereby reducing their workload and improving patrol efficiency. This study collected more than 2,000 images of real road pits containing complex conditions such as shadows and strong light in many places in Sichuan, and constructed a high-quality labeling data set. In the object detection task, compared YOLOv8 and Mask-R-CNN, select YOLOv8 as the basic network framework to identify potholes in on-board videos and capture images. Then, the three semantic segmentation models of YOLOv8, U-net and Mask-R-CNN were compared. Since Mask-R-CNN can better segment small potholes, it is selected as the segmentation model. The area is estimated and the inner diameter of the pothole is calculated using the mask to evaluate the warning level. After rigorous training and evaluation, the mAP50 of YOLOv8 for pits was 0.88 and the accuracy was 0.89. The frequency weight ratio of Mask-R-CNN in the semantic segmentation part was 0.55 and the ratio of 0.45. Experiments show that the system detection speed reaches 100FPS and the early warning accuracy rate exceeds 90%.

Keywords: Road Pits, Deep Learning, Object Detection, Convolutional Neural Network, Driving Recorder.

1. Introduction

This study focuses on the critical issue of road pothole detection. Traditional manual inspection is inefficient and often lags behind real-time conditions, making it difficult to discover and repair potholes promptly, which frequently leads to casualties and traffic accidents. Therefore, employing in-vehicle video analytics to identify road potholes rapidly and feed results back to authorities has become a core task in modern road maintenance.

Good road conditions are vital for socio-economic prosperity and travel safety. Historically, road inspection has evolved in three stages: (1) pre-20th-century manual checks with simple tools, (2) late-20th to early-21st-century automated sensor-based systems, and (3) current AI-driven methods. Deep-learning techniques—especially CNNs and Transformers—are now dominant. To improve robustness under varying weather, researchers also use GANs to synthesize training images under different illumination and climate conditions.

2. Literature Review

This study focuses on the critical issue of road pothole detection. Traditional manual road inspection methods are inefficient and suffer from numerous delays, making it difficult to detect and address potholes promptly, which often leads to injuries and traffic accidents. Therefore, utilizing vehicle-mounted video detection technology to efficiently identify and inspect road potholes, and rapidly transmitting the detection results to relevant regulatory authorities, has become a core task in current road inspection work.

Good road conditions hold significant importance in social life. They are not only a symbol of economic prosperity and social stability but also the foundation for ensuring the safety of travel. Historically, road inspection technologies both domestically and internationally can be roughly categorized into three stages. Before the 20th century, inspections mainly relied on manual checks assisted by simple tools; this approach was inefficient and prone to overlooking issues. From the late 20th century to the early 21st century, sensor-based automated detection technologies became mainstream, significantly improving inspection efficiency and accuracy. Currently, AI-based intelligent detection technologies are gradually becoming widespread, with deep learning technologies—especially convolutional neural networks (CNNs) and Transformer models—serving as core tools. Moreover, to address the impact of different weather conditions on detection results, researchers have also employed generative adversarial networks (GANs) to generate road images under various weather and lighting conditions, thereby effectively enhancing the generalization capability of detection models and enabling them to operate reliably in complex environments.

2.1. Global Research Progress

Internationally, progress is equally impressive. Malhar Khan [11] argues that accurate pothole detection is key to autonomous-driving safety and demonstrates that YOLOv8 greatly improves detection efficiency and lowers accident risk. Ohoud Alzamzami [12] employs UAV high-resolution imagery, allowing users to inspect specific road segments and select safer routes. Surya Sasank Ch [13] trains YOLOv8 to achieve 92 % mAP50 on potholes and uses spatial-resolution factors to estimate pothole dimensions for maintenance. Hoseini Mostafa [14] focuses on forest roads, using dash-cam data under various weather; his model achieves 0.58 tracking and 0.79 detection scores. Qian Li [15] shows that YOLOv5 maintains real-time, lossless processing while delivering high-accuracy pothole detection, supporting safe autonomous driving and road monitoring.

2.2. Domestic Research Status

Recent domestic work has achieved remarkable progress. In 2024, Liu Xu et al. [1] embedded dynamic Transformer feature-enhancement and dual-attention modules into YOLOv8, replacing its FPN with AFPN and markedly improving detection accuracy and speed. Liu Xu [2] also proposed an enhanced Cascade R-CNN for better pavement-defect detection. Xuan Yiguo [3] introduced deformable convolutions and a normalized Wasserstein-distance loss into YOLOv7, raising precision by 4.9 % and speed by 14 %. Wu Yanan [4] compared six ML models for pothole-depth prediction using the LTPP database and found XGBoost best, boosting TCN performance by 51.09 %. Zhu Huanxin [5] developed an improved Cycle-Dehaze algorithm for foggy driving scenes, outperforming others by 2.4 % PSNR and 3.5 % SSIM. Deng Hao [6] achieved lightweight pothole detection via data augmentation (flip, crop) for resource-limited scenarios. Dang Mingxia [7] collected custom datasets using smartphone videos. Zeng Shuai [8] integrated Coordinate Attention and SimCSPSPPF into YOLOv5 on RDD2022, significantly improving PA and mIoU. Xiang Guangde [9] combined attention modules with YOLOv5 to enhance small-defect feature extraction. Yu Lingjun [10] built a U-Net-based system that classifies and locates road damage, displaying GPS coordinates and maintenance advice on electronic maps.

3. Problem Statement

3.1. Production Process of precast Components

In recent years, the primary techniques for pothole repair in China have included radar detection technology, distributed fiber optic sensing technology, visual observation and measurement tools, UAV inspections, and infrared imaging technology. However, these technologies are costly to implement and are often affected by weather and road conditions. Manual inspections, on the other hand, require higher labor and time investment while offering low efficiency. With the development of artificial intelligence, this paper proposes a method that applies deep learning-based object detection algorithms to road pothole detection. For data collection, this paper suggests a simple approach that requires only a driver, a dashcam, and a vehicle to achieve real-time monitoring. This method significantly enhances the efficiency and coverage of data collection. Based on a large and effective dataset, the deep learning algorithm can quickly and accurately detect the location and shape characteristics of road potholes. It demonstrates strong robustness and anti-interference capabilities, enabling effective detection in various challenging conditions such as intense sunlight, rain, snow, and low-light environments, and it produces reliable detection results. This approach will greatly improve the efficiency of road pothole detection, provide robust data support for subsequent repairs, reduce the likelihood of traffic accidents caused by potholes, ensure the safety of passengers and property, lower maintenance costs, and generate significant economic benefits.

3.2. Method

3.2.1. Data Preprocessing Design

During the data preparation phase, a series of preprocessing operations were applied to the image data, mainly including image normalization, denoising, and format conversion. These steps effectively enhance data quality, providing a solid foundation for model training and subsequently optimizing the training outcomes. Specifically, for pavement pothole images, various data augmentation techniques such as rotation, flipping, scaling, and cropping were employed. These methods significantly increase data diversity, enrich the training samples, and effectively improve the model's generalization capability, enabling it to adapt better to different scenarios and conditions. Finally, the dataset, after preprocessing and data augmentation, was divided into training, validation, and test sets according to a predetermined ratio to meet the distinct requirements of model training, tuning, and performance evaluation. The data preprocessing workflow is illustrated in Figure 1.

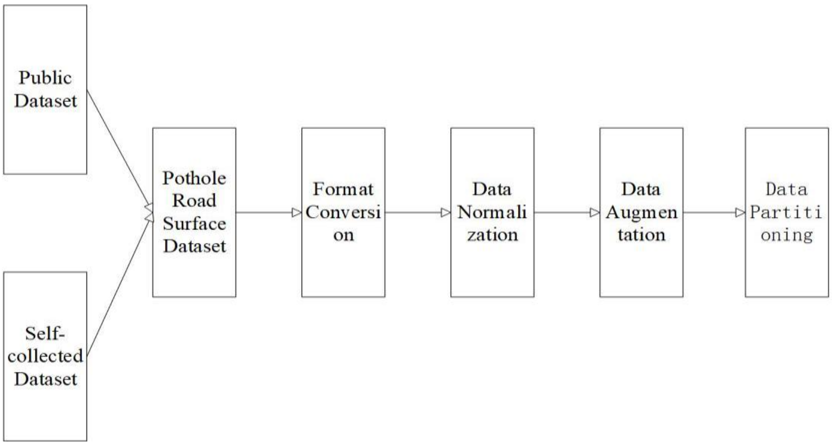

Figure 1. Data Processing Flowchart

This auxiliary diagnostic system is divided into three main modules: the road pothole target tracking module, the road pothole recognition and segmentation module, and the PYQT interface implementation module: Data Module: Both the pothole detection and pothole area calculation modules require data processing.

3.2.2. Module Design

Road Pothole Detection Implementation Module: This module is divided into target detection of road potholes and further recognition and segmentation of road potholes. With the accurate recognition provided by YOLOv8, further recognition and segmentation of pothole images are carried out using Mask-R-CNN. The area and internal diameter of potholes are calculated based on the number of masks. Key data such as loss values and accuracy are then collected to evaluate the model's performance. For tasks involving image segmentation, in addition to relying on these quantitative indicators, it is essential to visually inspect the segmentation results. Specifically, segmentation results can be displayed graphically to intuitively evaluate the model's segmentation accuracy, thereby providing a more comprehensive understanding of the model's practical effectiveness.

PYQT Interface Module: The PYQT interface is divided into two parts: interface management and the pothole early warning system. Potholes are rated according to their internal diameter: those with a diameter of less than 30 cm are classified as small potholes, those with a diameter of 30 cm to 50 cm are classified as medium potholes, and those with a diameter exceeding 50 cm are classified as large potholes.

3.2.3. Implementation Process of Object Detection and Recognition Segmentation Models

After the deep learning model is constructed, it is mainly divided into two parts: one is the object detection and tracking module, and the other is the recognition and segmentation module. the first model optimizes its parameters by learning from labeled samples. Specifically, the training process of deep learning typically follows the following steps:

(1) Data preprocessing: First, input data is fed into the YOLOv8 model and converted into a data format compatible with YOLOv8.

(2) Model Training: By adjusting parameters, the model is iterated multiple times to achieve the best detection performance.

Another model under the target recognition and segmentation module, it is mainly divided into the following two steps:

(3) Data Preprocessing: Perform denoising on the images processed in the previous step again, and standardize their format.

(4) Model Training: Initially, some images are used to pre-train the model to obtain a preliminary model. This model is then trained using the training dataset. After multiple rounds of training, the model begins to converge, ultimately resulting in a segmentation model. The segmentation model is then used to annotate the potholes in the images and to count the number of pixels contained in each annotated pothole. Based on the actual situation, a unit pixel area is set to calculate the area of each pothole.

3.2.4. Application and Design of Road Surface Pothole Detection and Segmentation

First, road inspectors need to import the video of the road section to be inspected into the system. During subsequent processing, each frame of the video is preprocessed to reduce the interference caused by weather conditions, video pixel quality, and non-pothole objects, allowing the system to focus more accurately on pothole areas. The preprocessed video frames are then input into the trained object detection model (YOLOv8) to identify road potholes. To ensure the efficiency of the entire process, the parameters identifying potholes are directly passed to the instance segmentation network. Using the Mask-R-CNN algorithm, the system can accurately calculate the pothole area and predict the pothole diameter. Based on the calculated and predicted results, a pothole with a diameter less than 0.3 meters is classified as a Level 3 pothole; if the diameter is greater than 0.3 meters but less than 0.5 meters, it is classified as a Level 2 pothole; and if the diameter exceeds 0.5 meters, it is classified as a Level 1 pothole. When road potholes reach Level 2 or Level 1, relevant personnel should take appropriate measures promptly.

This study, based on the PyQt5 framework and integrating the YOLOv8 and Mask-R-CNN models, has developed an efficient pothole detection application. Through carefully designed functional algorithms, the system can accurately predict pothole sizes and provide corresponding warning levels based on detection results. The system centers around a graphical user interface (GUI), providing users with an intuitive and easy-to-operate interactive platform, fully considering the characteristics and management requirements of road potholes, and aims to offer robust technological support for the management of road potholes.

In the system architecture, the interface primarily undertakes the tasks of presentation and interaction, encompassing three main functional modules: image detection for human-computer interaction, dashcam video detection, and real-time road monitoring. Images, dashcam videos, or real-time surveillance video captured by the user through the interface are transmitted to the backend model. The model processes these input data, generates functional response results, and provides real-time feedback to the interface. In addition, the system processes the response results through an intermediary layer, and when certain conditions are met, the results are transmitted to the interface for user viewing and decision-making.

4. Experiments and Results

4.1. Object Detection

4.1.1. Comparative Analysis of Experimental Results

This paper selects the Mask-R-CNN and YOLOv8 models, and uses pre-trained models to perform recognition on a video segment, comparing the two models in terms of potholedetection speed, precision, and mAP, as shown in Table 1.

Table 1. Comparison of Pothole Detection Model Results

	Speed/ms	Precision/%	Map/%
Mask-R-CNN	14.53	0.79	0.81
YOLOv8	4.89	0.89	0.88

Based on the comparison of the two models in the above table and considering the actual conditions of this experiment, real-time detection and computation of vehicle-mounted videos are required. Therefore, this study uses the YOLOv8 model to detect potholes.

4.1.2. Analysis of Object Detection Experiment Results

As can be seen from Figure 2, after 100 iterations of training, the first three charts indicate that YOLOv8 has shown significant improvements in both the detection and classification of potholes of various sizes. In the charts for val/box_loss, val/cls_loss, and val/dfl_loss, it is evident that the validation losses are similar to the training losses and both are decreasing, indicating that the model possesses good generalization capabilities. Furthermore, from the evaluation chart of mAP50-95, it is clearly observed that the model's accuracy has improved to a certain extent.

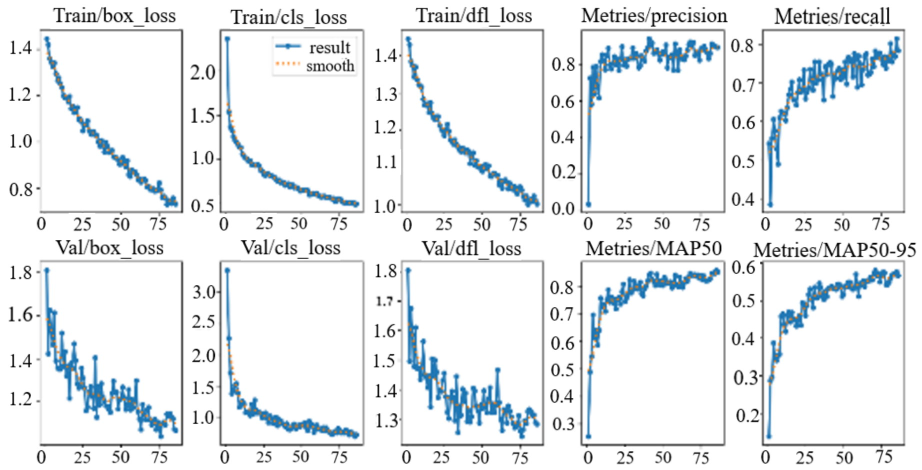

Figure 2. YOLOv8 Object Detection Evaluation Chart

In Figure 3.11, the model evaluation after 100 training iterations of YOLOv8 is presented.

From the F1-Confidence curve, it can be observed that the F1 score reaches its peak of 0.82 at a confidence level of 0.540. The second figure shows that precision achieves its highest point of 0.89 at a confidence level of 0.892. In the Precision-Recall curve, it can be seen that as recall increases, precision gradually decreases, indicating that improving recall may introduce more false positives, thereby reducing precision. In the Recall-Confidence curve, recall gradually decreases as confidence increases. This suggests that increasing confidence may cause the model to miss more positive samples, leading to a lower recall.

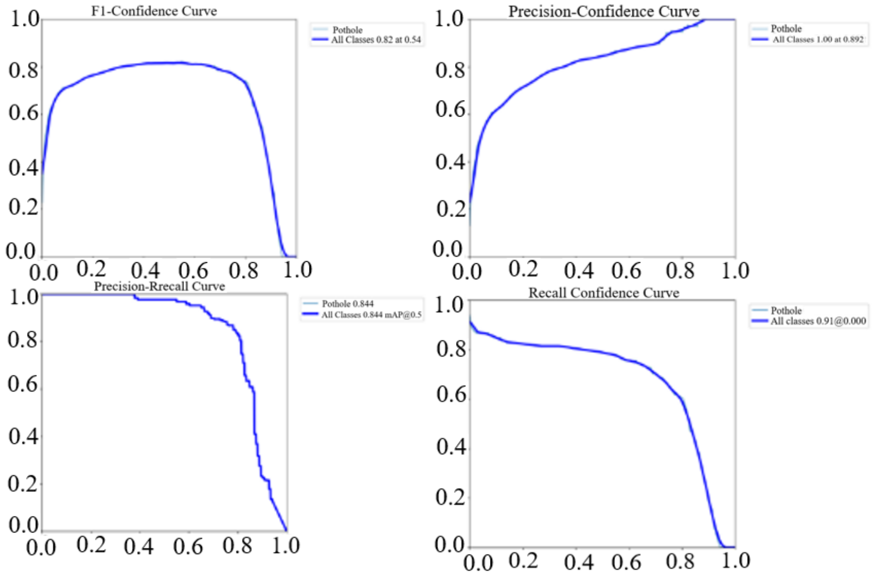

Figure 3. Confusion Matrix of YOLOv8

4.2. Segmentation Model

4.2.1. Principles of Computing

In this paper, with the aid of the mask information output by the Mask R-CNN model, the calculation of the target area is realized. Specifically, the calculation does not merely rely on the width and height of the mask, but rather infers the target area based on the distribution of pixels covered by the mask. By using the torch. sum() function, the number of pixels corresponding to the target within the mask is precisely counted, and the resulting value is denoted as total_area. In practical scenarios, each pixel corresponds to a fixed physical area, which we define as size. This value of size is not arbitrarily set but is determined by the resolution of the image as well as the specific proportional relationship between the actual scene and the image.

Based on the above definitions and statistical results, the actual area of the target, area, can be accurately obtained by multiplying the counted number of foreground pixels, total_area, by the physical area represented by each pixel, size. The specific calculation formula is:

$$\text{area}=\text{tatal\_area}\times\text{size}$$

When rating the pothole diameter (actual_length), this paper multiplies the physical area by the width of the pothole and then divides by the original width of the image, using the formula:

$$\text{actural\_lenth}=\frac{\text{width}\times\text{size\_value}}{\text{img\_width}}$$

4.2.2. Model Training Strategy

Based on training experiments with a series of object detection algorithms and an experimental dataset of annotated road potholes, training experiments were conducted on the YOLOv8 and Mask R-CNN models using the same learning rate, optimizer, and number of samples to ensure the accuracy and reliability of the experimental results. The efficiency and accuracy of these models in addressing the road pothole detection problem were evaluated and compared. The parameter settings for each model are shown in Table 2 below:

Table 2. Comparison of Model Training Parameters

	Image—size	LR	Optimizer	Batch Size	Epochs
YOLOv8	640*640	0.001	AdamW	32	100
Mask-R-CNN	640*640	0.001	AdamW	32	50

4.2.3. Analysis of Experimental Results Comparison

To achieve precise calculation of road pothole areas, a comparison was conducted among three commonly used models, and the optimal model was selected based on experimental evaluation metrics for further pothole recognition and segmentation. As shown in Table 3, among the three models, the Mask R-CNN model demonstrates superior and balanced evaluation metrics. Therefore, this study adopts the Mask R-CNN model for the pothole image segmentation stage.

Table 3. Comparison Table of Results from Two Segmentation Models

	PAcc	IoU	f- IoU
YOLOv8	0.87	0.40	0.93
Mask-R-CNN	0.94	0.55	0.45

From the evaluation charts of the three models above, it can be seen that Mask-R-CNN has a slight advantage in precision, accuracy, and recall. Therefore, in the next chapter on area calculation, this study adopts the Mask-R-CNN algorithm for research.

4.2.4. Analysis of Experimental Results

The learning rate is optimized through three stages: warm-up, cyclical adjustment, and final annealing, in order to enhance the training process. As shown in Figure 4, when the learning rate stabilizes at 6×10^-5, it has reached stability and no further adjustments are made.

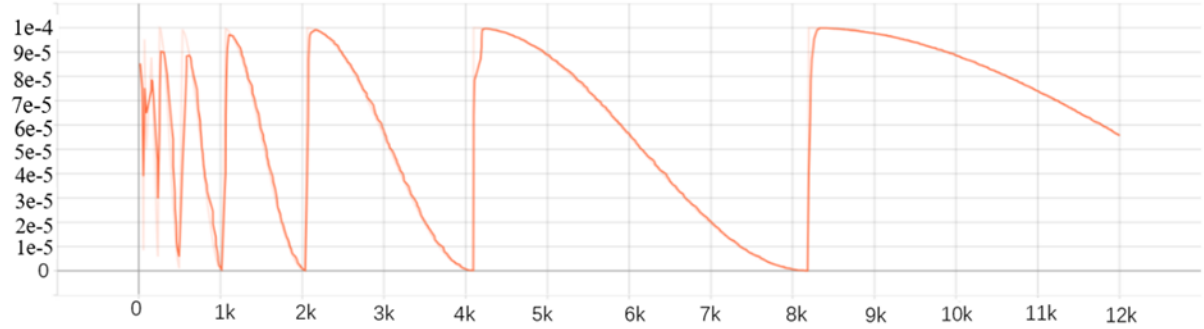

Figure 4. Mask-R-CNN Learning Rate

In the comparative experiment for road pothole calculation, the Mask-R-CNN model was used for training. In Figure 5, the x-axis represents the number of training iterations, and the y-axis represents the mean accuracy. It can be observed from the figure that the model has already converged after 50 training iterations, so training was halted to conserve resources. Additionally, when the model had been trained 32 times, the mean accuracy peaked at around 0.5. Therefore, in this experiment, the Mask-R-CNN model trained for 32 iterations was used as the mask computation model. After 32 training sessions, the Mask-R-CNN achieved an average precision of 0.5, accuracy of 0.91, recall of 0.82, F1 score of 0.84, and overall accuracy of 0.99.

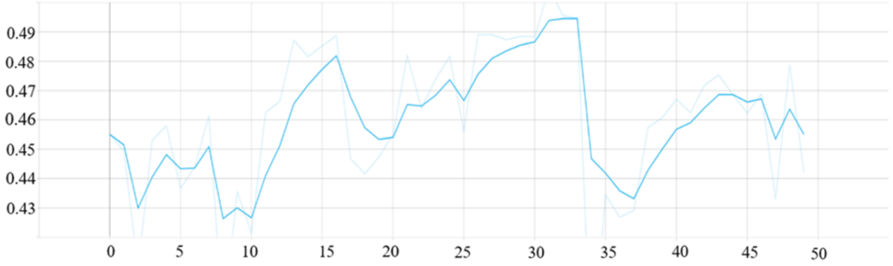

Figure 5. Map curve

4.3. Discussion

This study employs YOLOv8 and Mask R-CNN to enhance the precision and accuracy in calculating pothole areas and diameters. Compared to traditional methods, this system can provide faster and more accurate detection results, offering an intelligent and efficient solution for urban road management.

5. Conclusion

This paper proposes a road pothole detection system based on the YOLOv8 deep learning model. It combines dashcam data with existing algorithms to create a PyQt interactive interface and provides size-based grading and early warning for potholes according to their diameters. Experimental validation shows that the system performs excellently in terms of detection accuracy and real-time performance, achieving a mAP50 of 91% and a mAP50-95 of 66.2%, demonstrating high detection efficiency and accuracy.

Although this research has achieved certain results, there remains room for improvement. For instance, integrating a database that allows the upload of videos from dashcams would enable road inspectors to conveniently extract footage for road evaluation. This approach would not only increase the diversity of the data but also enhance the model's generalization capability.

Acknowledgments

This research was Supported by the "College Student

Innovation and Entrepreneurship Training Program Project (Project No.: S202410622087 and cx2025158).

References

[1] Liu Xu. Research on Road Surface Defect Detection Methods Based on Computer Vision and System Design [D]. Chongqing Jiaotong University, 2024.

[2] Liu Xu, Chen Lili, Fan Guodong, et al. Road Surface Defect Detection Based on an Improved Cascade R-CNN Algorithm [J]. Automation and Instrumentation, 2023, (10): 10-14.

[3] Xuan, Yiguo. Research on Pavement Crack and Pothole Damage Detection Based on Improved YOLO Algorithm [D]. Chongqing University of Technology, 2024.

[4] Wu Yanan, Wang Chao, Zhu Xingyi. Research on Pavement Structure Depth Prediction Based on Data Augmentation and Machine Learning [J/OL]. Journal of Southeast University (Natural Science Edition), 1-20 [2025-05-31].

[5] Zhu Huanxin. Research on Drivable Road Surface Segmentation Method for Intelligent Driving in Foggy Weather Based on Deep Learning [D]. Hubei University of Automotive Technology, 2024.

[6] Deng Hao. Research on Lightweight Pavement Pothole Detection Method Based on Deep Learning [D]. Harbin Institute of Technology, 2023.

[7] Dang Mingxia. Research on Monocular Road Pothole Detection Based on Computer Vision [D]. Chang'an University, 2024.

[8] Zeng Shuai. Semi-Supervised 3D Object Detection for Autonomous Driving [D]. Beijing University of Posts and Telecommunications, 2024.

[9] Xiang, Guangde. Research on Road Defect Detection and Identification Methods Based on Deep Learning [D]. Northeast Forestry University, 2024.

[10] Yu Lingjun. Research on Road Damage Detection System Based on Computer Vision [D]. Inner Mongolia University of Science and Technology, 2023.

[11] Khan M, Raza A M, Abbas G, et al.Pothole detection for autonomous vehicles using deep learning: a robust and efficient solution [J]. Frontiers in Built Environment, 2024,9.

[12] AlzamzamiO, BabourA, BaalawiW, etal. PDS-UAV: ADeepLearning-BasedPotholeDetectionSystemUsingUnmannedAerialVehicle Images [J]. Sustainability, 2024, 16(21): 9168-9168.

[13] ChSS, TallamT. Pothole detection and dimension estimation by deep learning [J]. IOP Conference Series: Earthand Environmental Science, 2024, 1326(1).

[14] MostafaH, StefanoP, StephanH, etal. Pothole detection in the woods: a deep learning approach for forest road surface monitoring with dashcams [J]. International Journal of Forest Engineering, 2024, 35(2):303-312.

[15] LiQ, ShiY, LiuQ, etal. Deep Learning-based Pothole Detection for Intelligent Transportation: AYOLOv5 Approach [J]. International Journal of Advanced Computer Scienceand Applications (IJACSA), 2023, 14(12).