Asphalt Pothole Detection via Grayscale-Texture Fusion and Fast R-CNN with Morph Postprocessing

Liang Guo[1*], Wei Han1, Haiming Cheng[1], Yang Ji[2]
[1]CCCC Third Highway Engineering Central China Construction Co., LTD
[2]The Second Construction Co., LTD. of China Construction First Group, Beijing, 100068, China
E-mail: liangguoo@outlook.com
[*]Corresponding author

Keywords: feature image recognition, pothole detection, fast R-CNN, neural network, neural networks, master element analysis, asphalt pavement

Received: March 13, 2025

Aiming at the problems of low recognition accuracy and insufficient feature extraction in existing vision-based pothole detection methods, this paper proposes a multi-stage detection method that integrates grayscale and texture features. The method flow consists of four stages: firstly, the pavement image is binarized to achieve pothole qualitative identification (92.2% accuracy) and coarse extraction based on shape features and local standard deviation; secondly, texture features of the candidate region are extracted through the grayscale covariance matrix and principal component analysis (PCA) is used to eliminate the feature redundancy; then, the texture region that conforms to the pothole lesion characteristics is aggregated with the results of the pothole lesion detection using fuzzy C-means clustering algorithm. characteristics of the texture regions are aggregated and spatially superimposed with the coarse extraction results; finally, the boundary is optimized by morphological post-processing to obtain accurate pothole segmentation results. On the dataset containing disturbing scenes such as cracks, gravels, water accumulation, etc., the method achieves a recall rate of 90.0% and a precision rate of 87.1%, in which 70.4% of the samples have an intersection and union ratio (IoU) of more than 80%, and 85.2% of the samples have an IoU of more than 70%. Experiments show that the detection performance of this method in complex pavement environments is significantly better than that of traditional single-feature detection models.

Povzetek: Prispevek obravnava večstopenjsko metodo zaznavanja udarnih jam v asfaltnih voziščih, ki združuje sivinske in teksturne značilke z globokim učenjem (Fast R-CNN). Predlagani pristop dosega visoko točnost in robustnost v kompleksnih okoljih.

1 Introduction

After the asphalt pavement is built and opened to traffic, with the increase in service life, coupled with the influence of extreme conditions such as rain, snow, high temperature, overloading, etc., there will be a variety of pavement potholes, such as cracks, rutting, congested packages, potholes, etc. The pavement maintenance work has become more and more important, and pavement inspection is an important preliminary stage of the maintenance work. Traditional manual detection requires a lot of labor and time, and automated detection has become a trend. Take the detection of cracks as an example, the common means of detection is to collect pavement images, with the help of computers, and the detection algorithm to complete the automatic identification and classification of cracks [1-3]. The increasing maturity of crack detection algorithms has promoted a series of automated crack repair machines [4-7]. However, potholes, as another common pavement pothole, the related automated detection and automated repair technology needs to be further improved. In this paper, the research on pothole automated detection technology is carried out to accurately recognize and extract potholes to support the evaluation of the degree of pavement pothole damage.

Currently, several common automated pothole detection methods include the vibration method [8-11], the 3D reconstruction method [12-17], and the vision-based detection method [18-21]. The vibration method uses acceleration sensors to sense the bumps and vibrations of the road surface to provide feedback signals to the vehicle to detect pothole damage, which is inexpensive equipment, occupies a small memory, and has a fast-processing speed. However, the method has the inherent disadvantages of misdetection and missed detection [22-23]. The three-dimensional reconstruction method uses three-dimensional technology to reconstruct the three-dimensional information of potholes, which also includes the three-dimensional laser scanning method [12-13], the stereo vision method [14-15], and the Kinect-based reconstruction method [16-17]. Among them, the 3D laser scanning method has

high detection accuracy, but the equipment is expensive and does not apply to universal pavement inspection. The stereo vision method is limited by the difficulty of matching images and a large number of calculations, which is difficult to apply in the actual pavement inspection environment. Kinect is a relatively new detection means in recent years, but it is still in the early exploration stage. The vision-based detection method uses a simple vehicle-mounted camera to collect pavement images, and can accurately detect potholes based on two-dimensional images. Comparing the above three pothole detection methods, it can be seen that the vision-based detection method is better than the vibration method in terms of detection accuracy, the price of the equipment is much lower than the three-dimensional detection method, and it can accurately detect potholes in pavements, which is easy to promote and use in the inspection of pothole damage in pavements and the evaluation of the degree of damage.

Vision-based pothole detection methods are analyzed using image processing techniques after capturing pavement images. Among them, Lin et al [18] used SVM, combined with texture information to train the pavement. After calculating the texture information of potential potholed areas, the SVM was used to recognize whether the potential areas belonged to potholes or not. The disadvantage is that the training samples have only texture features and no shape features, which can easily lead to cracks being misidentified as potholes, affecting the recognition and extraction accuracy. Buza et al [19] proposed to use spectral clustering to detect potholes in pavements. The disadvantage is that the algorithm only relies on the grayscale histogram data points for clustering, ignoring the texture information, which reduces the detection accuracy and extraction effect. Huidrom et al [20].

A heuristic decision logic approach combining area, standard deviation, roundness, and diameter information was used to detect and differentiate pavement cracks and potholes. Koch et al [21, 22] fit ellipse to the potholed area and combined the standard deviation of the area and the background to identify pavement potholes. Ryu et al [23-24] extracted potential pothole lesions by using the area and tightness to compute the potential lesion area and the histogram similarity and standard deviation of the background area to finally detect the presence of potholes in the pavement. These three detection methods focus on the identification of potholes, and all of them extract the potential potholed areas after binarizing the gray-scale images. Finally, it is determined whether the potential potholed area is a pothole or not. Since not all potholes have dark grayscale in the whole pothole region, the pothole region extracted after grayscale image binarization is usually only a part of the pothole, the potential pothole region extracted by these three methods based on grayscale image binarization is incomplete, which fundamentally determines the accuracy of the final extracted pothole region is not high [25-27].

Current vision-based pothole detection methods face challenges such as insufficient feature characterization and high false detection rate in complex environments (e.g., dense cracks, water-covered or uneven illumination), and the existing techniques mostly rely on single-modal features, making it difficult to balance the detection accuracy and real-time demand, which restricts their practical application in road inspection. In view of the above limitations, this study is dedicated to developing a multi-feature fusion collaborative detection framework, which focuses on solving the problems of missed detection and boundary blurring in complex interference scenarios through the joint analysis of grayscale geometric attributes and local texture statistics, combined with the hierarchical validation mechanism (coarse-grained filtering and fine-grained clustering superimposed on each other). The expected results include: 1) constructing a lightweight model architecture, adapting to edge computing platforms such as vehicle-mounted mobile devices, and realizing real-time detection and processing; 2) establishing a scalable framework for quantitative assessment of pavement potholes, providing theoretical support and technical paths for an engineered and time-sensitive road inspection system.

Therefore, this paper adopts a vision-based method to detect potholes [28-30]. Aiming at the shortcomings of existing vision-based pothole detection methods that cannot accurately extract potholes, this paper proposes an image texture-based pothole identification and extraction method, which can accurately identify and extract potholes, and helps to evaluate the degree of pavement pothole damage in terms of both number and area.

2 Image recognition method for pavement potholes

2.1 Image recognition algorithm

In this study, we propose a fully automated pothole quantification framework for asphalt pavements based on RGB sensors (see Fig. 1 for the process details), and its core process consists of four stages: ①multimodal data acquisition - synchronously acquiring pavement RGB images and depth point clouds; ②defect detection and localization The pothole bounding box is generated by Faster R-CNN, and the overlap detection is filtered by non-maximum suppression (NMS threshold=0.7); ③ Three-dimensional surface segmentation - the corresponding point cloud area is intercepted based on the bounding box coordinates, and the damage surface is extracted by the area growth algorithm (curvature threshold<0.05); ④ Damage quantification and analysis - the damage surface is quantified and analyzed. surface; ④ Damage quantization analysis - calculate the volume integral to projected area ratio of the segmented point cloud.The flow is shown in Figure 1.

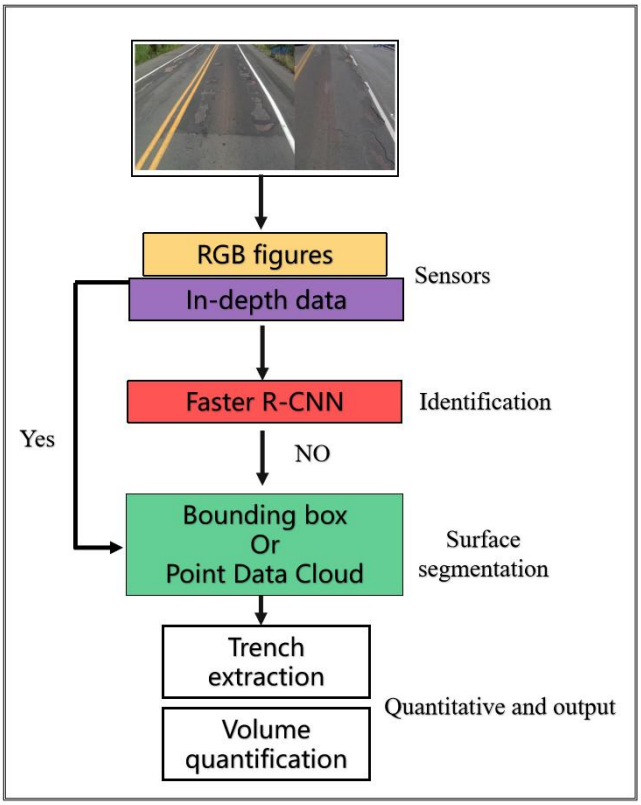

Figure 1: Process framework diagram

Microsoft Kinect V2 is a depth camera with the following parameters: an RGB camera of 1920?*80 pixels, a depth camera of 512* 24 pixels, a maximum depth distance of 4.5 m, and a minimum depth distance of 0.5 m. The resolution of the RGB image data and the point cloud data based on this camera is 1,920 *1,080 pixels and 512*24 pixels, respectively. The time-of-flight (ToF) concept can be realized using the RGB-D camera. It has both an IR emitter and an IR sensor. The IR emitter projects IR light onto the object, the light bounces off the surface and the IR sensor captures the reflected light. Since the distance between the IR sensor and the IR emitter is known, the 3D coordinates of each sensor pixel can be determined based on the time it takes for the IR light to propagate from the emitter to the sensor.

To detect and localize various asphalt pavement potholes, a Fast R-CNN-based detection method is used. This method only requires the use of RGB image data provided by the RGB-D sensor. Compared to R-CNN, Fast R-CNN uses a selective search method, which improves in terms of computational cost and running time. However, it is still unable to detect and localize targets in real-time. REN et al. proposed the Regional Candidate Network (RPN) to save the computational cost of localizing targets. The computational cost can be reduced by sharing the CNN architecture shown in Figure 2 and integrating the RPN into the existing Fast RCNN.

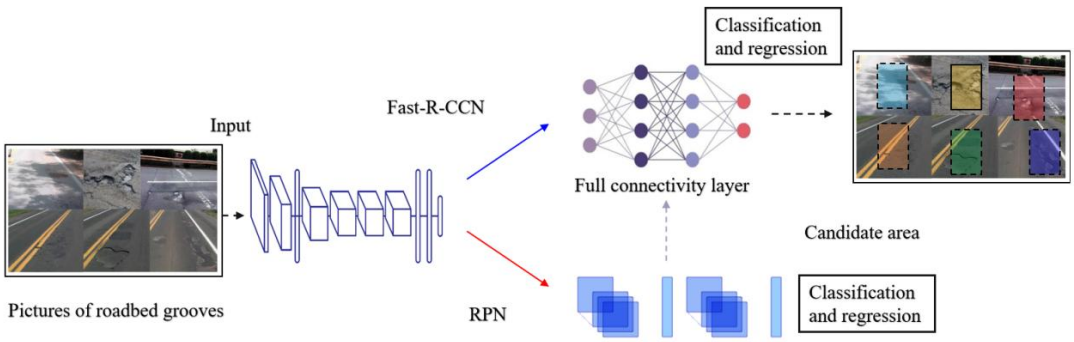

Figure 2: Process demonstration

The above algorithm is implemented in MATLAB environment, which mainly includes two processes, the first process is pavement pothole identification and initial extraction, and its main steps include: after noise reduction of the pavement image, the global threshold is utilized in the pothole. Combined with the standard deviation, the potholes are finally recognized and initially extracted. The second process is to cluster the pothole regions into one class using texture features and combine them with the initial extraction results obtained in the first process. The main steps include: calculating the texture feature vector using the grayscale covariance matrix and extracting the pothole region by fuzzy clustering after dimensionality reduction by principal element analysis. Superimposing the initial extraction results obtained from the first process, the potholes can be accurately extracted after morphological operations. Compared with existing vision-based pothole detection methods, the method in this paper can not only accurately recognize whether potholes exist in asphalt pavements, but also extract the shape of potholes more accurately. It helps to evaluate the degree of pothole damage from both the number and area of pavement potholes.

2.2 Pothole identification

This process is used for qualitative analysis and initial extraction of potholes to determine the presence or absence of potholes in the pavement image. The initial extraction of images with the presence of potholes is the preliminary work for the next step of accurate extraction. The main pipeline includes:

(1) RGB map is converted to grayscale;

(2) Image noise reduction. Median filtering is used to effectively smooth out impulse noise while protecting the sharp edges of the image;

(3) Binarization. Road potholes have a certain depth, and potholes inside the minor damage, in natural lighting conditions, the interior will appear in the shadow area or brightness of the darker areas, using binarization methods can be extracted from part of the pothole region, the binarization method used in this paper is the maximum interclass variance method;

(4) Morphological operations. After inverting the black and white pixels of the binary image, morphological erosion is used to eliminate fine cracks and small speckled regions. Morphological expansion is then utilized to fill the internal voids of the extracted pitted regions;

Statistical characterization of the shape of the connected domain. Calculate the ellipticity e, compactness C, and area A of the connected domain. ellipticity and compactness are defined below:

$$e = \frac{l_{\min}}{l_{\max}}$$
$$C = \frac{l_{\max}^2}{4\pi A} \quad (1)$$

Where: $l_{\max}$ denotes the length of the long axis of the ellipse having the same standard second-order central moment as the region; $l_{\min}$ denotes the length of the short axis of the ellipse having the same standard second-order central moment as the region; and A denotes the area of the connected region.

Based on Eq. (3), the connected area in the picture is identified as a potential pothole area.

$$R = \begin{cases} \text{no} & \text{if} — A > 0.8 \times A_T \\ \text{yes} & \text{else if } (e > T_e) \&\& (C > T_C) \&\&\& (A > T_A) \end{cases} \quad (2)$$

Where, A denotes the area of the connected domain .A_T denotes the total area of the image. T is the threshold of a region. e is the ellipticity .C is the compactness.

The role of the first discriminant condition in Eq. (2) is to eliminate the influence of the road marking, when the road image exists, the areas other than the marking are extracted by the binarization process and meet the second discriminant condition in Eq. This will lead to the selected potential pothole area R. It is too large, which makes the computation of the standard deviation in step f) increase dramatically. Also, since the location of the pavement markings has few through traffic and generally does not have potholes, the effect of the markings can be eliminated and the computing speed can be improved by comparing the total area A of the connectivity domain with the total area Ata of the image. The role of the second discriminant condition is to filter out the cracks and small spots remaining after the morphological operations.

(5) Calculate the standard deviation STD of the gray-scale image corresponding to the potential pothole region.

A suitable threshold TsTD is selected to discriminate whether the connected domain R belongs to the pothole region based on equation (3). Where the threshold Tsrp is also determined by trial calculation of 50 pavement images as described in Section 2.1.

$$R = \begin{cases} \text{no} & \text{if} — \text{STD} > T_{\text{STD}} \\ \text{yes} & \text{otherwise} \end{cases} \quad (3)$$

Throughout the recognition process, step e) calculates shape information, which serves to eliminate the effects of markers, cracks, small patches, etc. Areas that are potholes in shape are selected. The influence of patches, shadows, and large pieces of oil can be further eliminated by the judgment of standard deviation STD in step f). The presence of potholes in the pavement image is finally determined. This calculates the shape and texture information sequentially, which can improve computational efficiency. The detection results of pothole identification are evaluated based on equation (4).

$$\begin{aligned} \text{recall} &= TP/(TP+FN) \\ \text{precision} &= TP/(TP+FP) \\ \text{accuracy} &= (TP+TN)/(TP+FP+TN+FN) \end{aligned} \quad (4)$$

where: tp stands for the sample images that are correctly detected as potholes. tn stands for the sample images that are correctly detected as non-potholes. fp stands for the sample images that are incorrectly detected as potholes. fn stands for the sample images that are incorrectly detected as non-potholes. recall is the proportion of recalled samples (TP + FN). precision is the accuracy rate, which stands for the proportion of all correctly retrieved samples (TP + FP) to all samples retrieved (TP + FP). retrieved (TP) as a proportion of all samples that should have been retrieved (TP) as a proportion of all samples that were retrieved (TP + FP). accuracy is the precision rate, representing the ratio of all correctly classified samples (TP + TN) to the total number of samples (TP + FP + TN + FN).

Combining texture clustering and morphological processing, the pothole region can be extracted more accurately, and main pipeline is as follows:

(1) Extraction of pothole texture information.

The image is divided into 5 x 5 sub-blocks, and the contrast, correlation, energy, and homogeneity of each sub-block in the 0 to 135-degree direction are counted using the grayscale covariance matrix [24-25].

(2) Texture feature vector dimensionality reduction.

The texture feature vector calculated in step b) is up to 16 dimensions, and the redundant information is eliminated by principal component analysis. Generally, the eigenvectors corresponding to the eigenvalues with a cumulative contribution rate of 85%~95% or more are selected as the principal component transform vectors after dimensionality reduction. In this paper, the feature vector is reduced to 5 dimensions based on the calculation results, as shown in Figure 2(b).

(3) Fuzzy C-mean clustering.

The texture feature vectors are clustered using the FCM clustering algorithm, which labels the background region pixels as 0 and the pit and groove region pixels as 1.

Superimpose the initial extraction results from Section 2.1 and perform morphological operations. The purpose of superimposition is to combine the grayscale extraction results with the texture extraction results. The superimposed image needs to be subjected to morphological operations, starting with an open operation to eliminate the influence of fine regions and calculate the connected domain with the largest area. Then the closed operation is utilized to fill the internal voids of the pits. Finally, morphological corrosion is utilized to improve the edge detection accuracy.

(4) The Canny algorithm is used to extract the pothole edges.

Figure 3 represents several main processes of pit and groove extraction. From Figure 3(a) it can be seen that the texture feature vector is reduced to 5 dimensions. The cumulative contribution of the feature values can reach more than 85%. From Figure 3(b) it can be seen that due to the inconspicuous texture of the dark areas inside the pit troughs and the presence of rough textures and cracks on the outside of the pit troughs, the pit trough areas after FCM clustering are not coherent and there are cluttered white pixels on the outside. From Figure 3(c), it can be seen that combined with the initial extraction results, the pothole region can be well segmented after morphological operations, and the extracted pothole edges are shown in Figure 3(d).

The final extraction results are evaluated using overlap degree and ratio. Wherein, the extraction results of a single pothole are evaluated using the overlap degree, i.e., the degree to which the extracted pothole region overlaps with the original image pothole region. The extraction results of multiple potholes are evaluated using ratio, i.e., the ratio of the number of images with a particular degree of overlap to the total number of images.

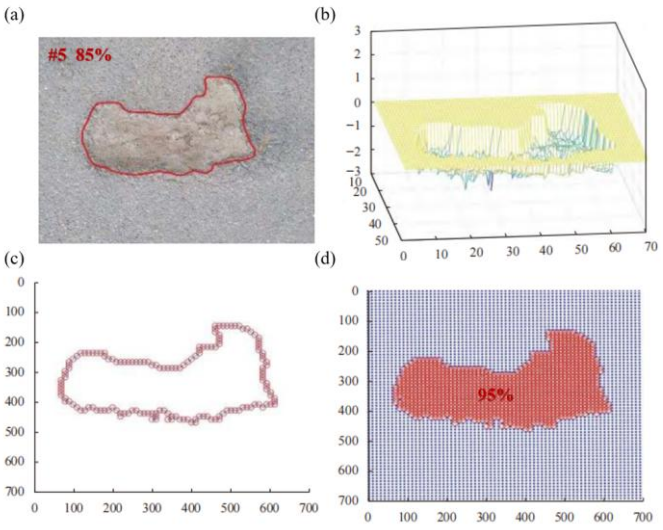

Figure 3: Pothole pit identification process and results

3 Experimental process and result analysis

3.1 Pavement pothole identification

This study dataset contains 140 pavement images (480 × 640 pixels, resolution 1.77 mm²/pixel) screened from Google Open Images V7, covering five categories of defects: cracks, patches, oil, shadows, and potholes, with sample sizes of 35, 26, 23, 29, and 17 for each category, respectively, in order to reflect the actual road maintenance in the Natural distribution differences. The data covers five types of scenarios, including urban roads, bridges and tunnels, and includes complex lighting conditions such as bright light, cloudy days, nighttime (15% of the test set), and rain and fog (10%). The labeling process was completed by three road engineers using the LabelMe tool to complete the pixel-level semantic segmentation, and finally generated high-precision ground truth labels through three stages of quality control.50 of these images are used for manual training to select appropriate thresholds for the qualitative identification of potholes in the pavement and the remaining 90 images are used to test the accuracy of the algorithms in this paper.

The training set and validation set are different sample sets, and the two are not duplicated. According to the pavement pothole recognition method proposed in Section 2.2, the key lies in the selection of thresholds Te, TA, TC., and TSTD, where the threshold Te is selected concerning Ryu's research results [23-24]. Based on the actual size of the pixel and the minimum diameter of the pit (150 mm), T is calculated to be about 10,000 pixels. The remaining two thresholds were determined by trial calculations on 50 pavement images. Based on equation (4) in section 2.2, the trial calculation results are statistically plotted to show the trends of recall, precision, and accuracy with the thresholds Te, TSTD, as shown in Figure 3. From Figure 4 it can be seen that as the threshold Te increases, recall shows a decreasing trend, precision shows an increasing trend, while accuracy shows an increasing and then decreasing trend. When Te = 0.14, both accuracy and recall are at their maximum values, but the value of precision is smaller. When Te increases to 0.16, recall, precision, and accuracy are all greater than 0.82, which is more desirable. If Te is increased further, although precision will become larger, recall and accuracy will become smaller rapidly, so it is more reasonable to set the value of Te at 0.16. Similarly, the value of TSTD is 0.12. Where the detection region makes the entire picture interval.

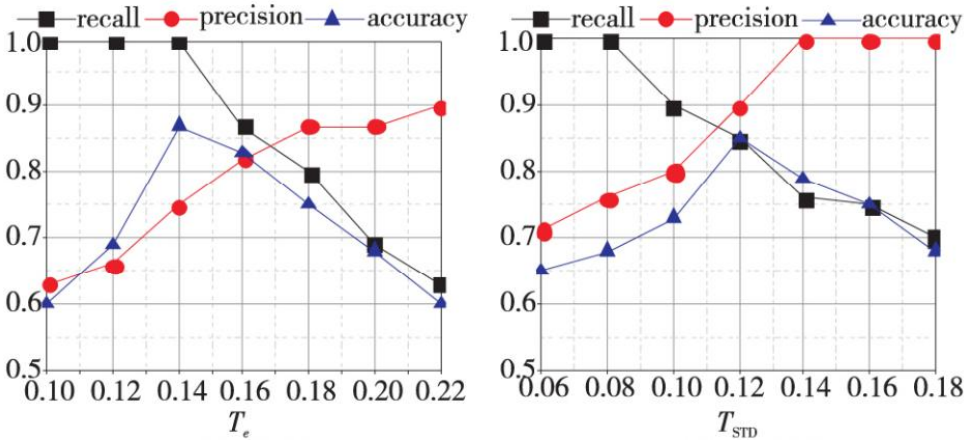

Figure 4: Evaluation comparison of recognition results

Combined with the above selection of thresholds, the specific values of the four parameters Te, TA, T., and TsrD are shown in Table 1.

Table 1: Parameter thresholds for slurry pit identification

Thresholds	Retrieve a value	Thresholds	Retrieve a value
Te	0.16	TA	10000
TC	0.05	TSTD	0.12

After the thresholds were selected, the algorithm of this paper was tested with the remaining 90 pavement images, and 25 pavement images with transverse longitudinal or mesh cracks. 8 images of good pavement. Pavement images containing patches 12 sheets. 5 sheets with spots and oil stains. 30 sheets with potholes. 10 sheets with shadows. The recognition effect of the algorithm is calculated based on equation (3,4) and is shown in Table 2. When the number of clusters (K=5), the DBI is 0.58, the SC is 0.73, and the silhouette coefficients of all the subclasses are >0.6. The results show the reasonableness of the clustering results in this study.

Table 2: Statistics of pothole identification for different types of pavement images

	Favorable	Fissures	Patch (for mending clothes, tires, etc)	Greasy and dirty	Low definition	Pothole	Aggregate
TP	/	/	/	/	/	/	27
FP	0	0	2	0	2	-	4
TN	8	25	10	5	8	-	56
FN	/	/	/	/	/	/	3
Reca.	/	/	/	/	/	/	90.0%
Proc.	/	/	/	/	/	/	87.1%
Accu.	/	/	/		/ /	/	92.2%

From the statistical results in Table 3, it can be seen that some patches and shadows were misidentified as potholes, mainly because the rough texture of the detected area led to a large standard deviation. Some potholes were misidentified as pavements without pothole, mainly due to the presence of dust or water inside these potholes, which made the potholes have large internal gray values and obscure texture features, resulting in the pothole region not being extracted during thresholding. Or the pothole region was extracted with a small standard deviation. Overall, however, the recall rate for recognizing the presence or absence of potholes in the pavement was 90.0%, the precision rate was 87.1%, and the accuracy rate was 92.2%. In comparison, as described in Table 3, Ryu's method has a recall rate of 73.3%, a precision rate of 80.0%, and an accuracy rate of 73.5% for recognizing potholes in pavements, indicating that this paper's method has a high degree of reliability in recognizing potholes.

Table 3: Comparison of pothole pit identification by different methods (%)

Methodologies	Reca.	Proc.	Accu.
Ryu	73.3	80.0	73.5
the main body of a book	90.0	87.1	92.2

Table 4 shows the field test results of the established method and the method in this paper. Existing pothole detection methods are generally limited by single feature extraction (e.g., SVM relies on grayscale, spectral clustering only uses texture) or simple scene adaptation (e.g., heuristic morphology has less than 73.8% recall under complex interference), resulting in significant performance degradation on highly disturbed datasets (IoU>70% sample share is generally below 65%). In contrast, this paper achieves multi-stage synergistic optimization through grayscale-texture feature fusion (combining shape, local standard deviation and PCA-optimized texture features) and hierarchical processing architecture (coarse extraction for qualitative recognition with fuzzy C-means + Fast R-CNN fine clustering superimposed). Morphological post-processing is further introduced to correct the clustering boundary fuzzy problem, and finally 90.0% recall and 87.1% precision are achieved in high interference scenarios, and the percentage of IoU>70% samples is increased to 85.2%, which is 14.8% higher than the optimal baseline method, verifying the significant advantages of multi-feature fusion and hierarchical strategy in complex pavement detection.

Table 4: Performance comparison of existing pothole detection methods

Method Category	Feature Type	Dataset Complexity	Accuracy (%)	Recall (%)	Percentage of Samples with IoU > 70%
SVM	Single grayscale feature	simple scenarios	72.3	68.5	52.1
Spectral clustering	Texture features	Moderate interference	79.6	75.2	63.8
Heuristic morphology	Shape features	simple scenarios	81.4	73.8	58.9
Methods in this paper	Gray-scale-texture fusion	High interference scenarios	87.1	90.0	85.2

3.2 Pavement pothole extraction

After the potholes are recognized, the initial extraction results of the potholes can be obtained, and the final extraction of the 27 pavement images detected to contain potholes is performed according to the method in Section 1.2, combined with the extraction results of the texture features. To analyze the advantages and shortcomings of this paper's algorithm, Figure 5 shows the results of this paper's method for recognizing and extracting several typical potholes. The first column (a) is the pavement image to be detected, the second column (b) is the result of pothole identification, and the third column (c) is the result of pothole extraction.

The grayscale features of the pothole shown in Figure 5-1(a) are obvious, and there is little difference between the pothole identification results and the extraction results. The grayscale features of the potholes shown in Figure 5-2(a) are missing. After combining the texture features, the extracted pothole region is more accurate. The pothole shown in Figure 5-3(a) contains water inside, and both gray-scale features and texture features are missing, which affects the initial extraction based on gray-scale and texture feature clustering. Figure 5-3 (a) shows a small pothole with mild breakage, and its extraction result 5-3 (c) shows that this paper's method can also obtain good results for the extraction of mild potholes. From the overall comparative analysis in Figure 5, it can be seen that this paper's method combines pothole grayscale and texture features, which applies to both heavy potholes and light potholes and can obtain accurate pothole identification and extraction results.

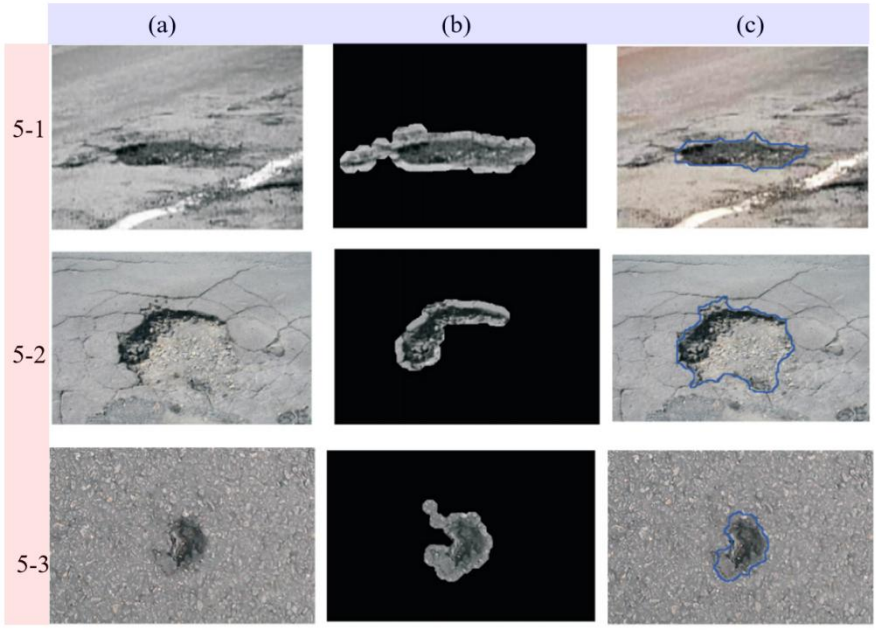

Figure 5: Recognition and extraction results of several typical sinkholes

Table 5 Statistics of the extraction results of 27 pavement images, the number of extracted pothole regions with more than 90% overlap with the original image pothole regions is 48.1% of the total number of images. The ratio of images with an overlap of 80% or more is 70.4%. The ratio of those with 70% or more is 85.2%. The remaining pothole images were detected with low overlap. The main reason is the presence of dust, water, or unbroken boulders inside the potholes, which results in the lack of both texture features and gray-scale features. The extraction results of 30 potholes in Ryu's paper are statistically calculated and compared with the extraction results of this paper's method, as shown in Table 6, which shows that this paper's method is more accurate in pothole extraction.

Table 5: Overlap statistics of pothole images

Overlap	Ratios	Overlap	Ratios	
>90%	48.1%	>70%	85.2%	
>80%	70.4%	<70%	14.8%	

Table 6: Comparison of the results of slot pit extraction by different methods (%)

Ratios	>90%	>80%	>70%	<70%
This paper	48.1%	70.4%	85.2%	14.8%
Ryu	29.6%	40.0%	59.3%	40.7%

Comparison experiments between this paper's method and existing mainstream methods (e.g., the heuristic morphological algorithm proposed by Ryu et al.) show that, in the complex scenario with dense cracks and serious waterlogging interference, the recall rate of this paper's method (90.0%) is improved by 16.2% compared with that of Ryu's method (73.8%), and the percentage of samples with IoU>70% is increased from 58.9% to 85.2%. This advantage stems from the following mechanisms: 1) the Ryu method relies only on morphological closure operations and fixed threshold segmentation, which is sensitive to texture noise (e.g., gravel artifacts) and leads to over-segmentation; whereas in this paper, we can effectively differentiate between real potholes and interferences through the synergy of grayscale-texture features (local standard deviation filtering + fuzzy C-means clustering); 2) in the scenes with uneven high illumination or partially water-covered potholes, the Ryu method is prone to miss detection due to the lack of multi-stage verification, while the superposition strategy of coarse extraction and fine clustering in this paper reduces the miss detection rate through spatial consistency constraints. However, in extreme low-resolution images (<0.5m/pixel), the accuracy of this method may decrease by about 5% due to the degradation of texture features.

4 Conclusion

Aiming at the shortcomings of existing vision-based pothole detection methods that focus only on recognition and have low extraction accuracy. In this paper, the method combines texture feature clustering based on the initial extraction of gray-scale features to achieve good pothole recognition and extraction results. It helps to comprehensively evaluate the degree of pavement pothole damage from both quantity and area.

The recognition process uses ellipticity, compactness, and area to screen out potential pothole regions, and uses the standard deviation of the regions to determine whether the potential pothole regions truly contain potholes. The recall rate is 90.0%, the precision rate is 87.1% and the accuracy rate is 92.2%, which indicates that the combination of shape features and standard deviation can achieve better pothole qualitative identification results.

The extraction process combines the results of texture clustering and the initial extraction results based on gray-scale features. The ratio of the number of extracted potholes to the number of images with an overlap of 80% or more of the original image potholes is 70.4%, and the ratio of the overlap of 70% or more is 85.2%. This result shows the good effect of the combined use of grayscale and texture features for quantitative extraction.

The experimental results show that as long as the texture or grayscale features of the pothole are significant, even if the edge of the pothole contains debris, cracks, or the interior contains part of the standing water, the qualitative identification and quantitative extraction can be carried out more accurately. On the contrary, the simultaneous absence of grayscale and texture features will seriously affect the accuracy of the algorithm, and how to solve this problem is a future research direction of this paper.

Appendix

```
# PyTorch-based Fast R-CNN core implementation (simplified)
import torchvision
from torchvision.models.detection.import FasterRCNN
from torchvision.models.detection.rpn import AnchorGenerator

# 1. Model construction
backbone = torchvision.models.resnet50(weights=“IMAGENET1K_V2”)
backbone.out_channels = 2048 # ResNet-50 backbone network
anchor_generator = AnchorGenerator(sizes=((32, 64, 128),), aspect_ratios=((0.5, 1.0, 2.0),))
model = FasterRCNN(
    backbone.
    num_classes=6, # 5 classes of defects + background
    rpn_anchor_generator=anchor_generator,
    box_score_thresh=0.8 # confidence threshold for inference phase
)

# 2. Hyperparameter settings
optimizer = torch.optim.AdamW(model.parameters(), lr=3e-4, weight_decay=1e-4)
lr_scheduler = torch.optim.lr_scheduler.StepLR(optimizer, step_size=5, gamma=0.1)
loss_fn = model.get_loss() # default multitask loss (classification + regression)

# 3. Training process
for epoch in range(20): # total training cycles
    for images, targets in train_loader: # data loading (with data augmentation)
        optimizer.zero_grad()
```

```
            loss_dict = model(images, targets) # forward propagation
            total_loss = sum(loss for loss in loss_dict.values()) # total loss
            total_loss.backward() # backward propagation
            optimizer.step()
        lr_scheduler.step()
```

Funding

Key R&D and Promotion Special Project (Science and Technology Research) in Henan Province, Research on key technologies of road damage detection based on deep learning (232102210108).

References

[1] Peng B, Jiang Yangsheng, Han Shifan, et al. A review of automatic recognition algorithms for pavement crack images. Highway Transportation Science and Technology, 2014, 31(7):19-25. https://doi.org/10.1061/JHTRCQ.000043.

[2] Jing Rong, Yuli Pan. Image-based crack recognition method for cement grooved pavement. Highway Transportation Science and Technology, 2012, 29(3): 45-50.

[3] Lv Yan, Qu Shiru. SIFT-based pavement crack alignment and splicing algorithm. Highway Transportation Science and Technology, 2012, 29(2): 23-28.

[4] Kim Y S, Yoo H S, Lee J H, et al. Chronological development history of X-Y table-based pavement crack sealers and research findings for practical use in the field. Automation in Construction, 2009, 18(6):513-524. https://doi.org/10.1016/j.autcon.2009.02.007.

[5] Lee J H, Yoo H S, Kim Y S, et al. The development of a machine vi-sion-assisted teleoperated pavement crack sealer. Automation in Construction, 2006, 15 (5):616-626. https://doi.org/10.22260/ISARC2004/0028.

[6] Tsai Y J, Kaul V, Yezzi A. Automating the crack map detection process for machine operated crack sealer.Automation in Construction, 2013, 31(10):10- 18. https://doi.org/10.1016/J.AUTCON.2012.11.033.

[7] Kim T, Ryu S K. Review and analysis of pothole detection methods. Journal of Emerging Trends in Computing and Information Sciences, 2014, 5(8):603-608.

[8] Yu B X, Yu Xinbao.Vibration-based system for pavement condition evaluation[C]//Proc of the 9th International Conference on Applications of Advanced Technology in Transportation. Reston, VA: American Society of Civil Engineers, 2006:183-189. https://doi.org/10.1061/40799(213)3.

[9] Zoysa K D, Keppitiyagama C, Seneviratne G P, et al. A public trans-port system-based sensor network for road surface condition monitoring[C]//Proc of Workshop on Networked Systems for Developing Regions.New York: ACM Press, 2007: Article No.9. https://doi.org/10.1145/1326571.132658.

[10] Erikson J, Girod L, Hull B, et al. The pothole patrol: using a mobile sensor network for road surface monitoring[C]//Proc of the 6th International Conference on Mobile Systems, Applications, and Services. New York: ACM Press, 2008:29-39. https://doi.org/10.1145/1378600.1378605.

[11] Rode S S, Vijay S, Goyal P, et al. Pothole detection and warning system [C]//Proc of International Conference on Electronic Computer Technology. Washington DC: IEEE Computer Society, 2009:286-290.10.1109/ICECT.2009.152.

[12] Chang KT, Chang J R, Liu J K. Detection of pavement distress using 3D laser scanning technology[C]//Proe of ASCE International Conference on Computing in Civil Engineering. Reston, VA: American Society of Civil Engineers, 2005: 1-11. https://doi.org/10.1061/40794(179)103.

[13] Li Qingguang, Yao Ming, Yao Xun, et al. A real-time 3D scanning system for pavement distortion inspection.Measurement Science and Technology, 2010. 21(1):15702-15709. DOI 10.1088/0957-0233/21/1/015702.

[14] Wang KC P. Challenges and feasibility for a comprehensive automated survey of pavement conditions [C] //Proc of the 8th International Conference on Applications of Advanced Technologies in Transportation Engineering. Reston, VA: American Society of Civil Engineers,2004:531-536. https://doi.org/10.1061/40730(144)9.

[15] Hou Zhiqiong, Wang K C P, Gong Weiguo. Experimentation of 3D pavement imaging through stereovision [C]//Proc of International Conference on Transportation Engineering. Reston, VA: American Society of Civil Engineers, 2007:376-381. https://doi.org/10.1061/40932(246)6.

[16] Joubert D, Tyatyantsi A, Mphahlehle J, et al. Pothole tagging system. Proc of the 4th Robotics and Mechatronics Conference of South Africa. Pretoria: CSIR,2011:1-4.

[17] Moazzam I, Kamal K, Mathavan S, et al. Metrology and visualization of potholes using the Microsoft Kinect sensor [C] //Proc of the 16th International IEEE Annual Conference on Intelligent Transportation Systems. Piscataway, NJ: IEEE Press,2013:1284-1291. 10.1109/ITSC.2013.6728408.

[18] Lin Jin, Liu Yayu. Potholes detection based on SVM in the pavement distress image [C] //Proc of the 9th International Symposium on Distributed Computing Washington DC: IEEE Computer Society, 2010:544-547. DOI: 10.1109/DCABES.2010.115.

[19] Buza E, Omanovic S, Huseinnovic A. Pothole detection with image processing and spectral clustering [C] //Proc of the 2nd International Conference on Information Technology and Computer Networks.2013:48-53.

[20] Huidrom L, Das L K, Sud S K. Method for automated assessment of potholes, cracks, and patches from road surface video clips. Proce-dia-Social and Behavioral Sciences, 2013, 104 (3):312-321. https://doi.org/10.1016/j.sbspro.2013.11.124.

[21] Koch C, Brilakis I. Pothole detection in asphalt pavement images. Advanced Engineering Informatics,2011,25 (3):507-515. https://doi.org/10.1016/j.aei.2011.01.002.

[22] Koch C, Jog G M, Brilakis I. Automated pothole distress assessment using asphalt pavement video data [J]. Journal of Computing in Civil Engineering,2013,27 (4):370-378. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000232.

[23] Ryu S K, Kim T, Kim Y R. Image-based pothole detection system for ITS service and road management system [J]. Mathematical Problems in Engineering,2015,2015 (9): Article ID 968361. https://doi.org/10.1155/2015/968361.

[24] Chang Chen Hao, Liu Rufei, Chai YN, et al. A pavement pothole extraction method for point cloud profile characterization, Geospatial Information, 2021, 19(2): 9-13. [doi:10.3969/j.issn.1672-4623.2021.02.003]

[25] NOH Y, KOO D, KANG Y M, Et al. Automatic crack detection on concrete images using segmentation via fuzzy C-means clustering [C] //2017 International Conference on Applied System Innovation (ICASI). ie, 2017: 877-880. Doi: 10.1109/Icasi.2017.7988574.

[26] Cha Y J, Choi W, Buyukozturk O. Deep learning-based crack damage detection using convolutional neural networks [J]. Computer-Aided Civil and Infrastructure Engineering, 2017, 32(5): 361 -378. https://doi.org/10.1111/mice.12263.

[27] Kamal K, Mathavan S, Zafar T, et al. Performance assessment of Kinect as a sensor for pothole imaging and metrology. International Journal of Pavement Engineering, 2018, 19(7): 565-576. https://doi.org/10.1080/10298436.2016.1187730.

[28] Torok M M, Golparvar-Fard M, Kochersberger Kb. Image-based automated 3D crack detection for post-disaster building assessment. Journal of Computing in Civil Engineering, 2014, 28 (5): A4014004. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000334.

[29] Tadas Žvirblis, Armantas Pikšrys, Damian Bzinkowski, Mirosław Rucki, Artūras Kilikevičius, Olga Kurasova, Data Augmentation for Classification of Multi-Domain Tension Signals, Informatica ,2024, 35, 883-908. https://doi.org/10.15388/24-INFOR578.

[30] Isah AD, Aibinu A M, Olaniyi O M, et al. Development of asphalt paved road pothole detection system using modified color space approach[J]. Journal of Computer Science and Its Application, 2018, 25(2). https://doi.org/10.4314/jcsia.v25i2.