

Research Article

Research on Road Adhesion Condition Identification Based on an Improved ALEXNet Model

QiMing Wang ¹, JinMing Xu,¹ Tao Sun,¹ ZhiChao Lv,² and GaoQiang Zong¹

¹School of Mechanical Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China

²School of Automotive Studies, Tongji University, Shanghai 201804, China

Correspondence should be addressed to QiMing Wang; wang.qiming2008@163.com

Received 5 February 2021; Revised 10 June 2021; Accepted 28 June 2021; Published 17 July 2021

Academic Editor: Wen LIU

Copyright © 2021 QiMing Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Automotive intelligence has become a revolutionary trend in automotive technology. Complex road driving conditions directly affect driving safety and comfort. Therefore, by improving the recognition accuracy of road type or road adhesion coefficient, the ability of vehicles to perceive the surrounding environment will be enhanced. This will further contribute to vehicle intelligence. In this paper, considering that the process of manually extracting image features is complicated and that the extraction method is random for everyone, road surface condition identification method based on an improved ALEXNet model, namely, the road surface recognition model (RSRM), is proposed. First, the ALEXNet network model is pretrained on the ImageNet dataset offline. Second, the weights of the shallow network structure after training, including the convolutional layer, are saved and migrated to the proposed model. In addition, the fully connected layer fixed to the shallow network is replaced by 2 to 3, which improves the training accuracy and shortens the training time. Finally, the traditional machine learning and improved ALEXNet model are compared, focusing on adaptability, prediction output, and error performance, among others. The results show that the accuracy of the proposed model is better than that of the traditional machine learning method by 10% and the ALEXNet model by 3%, and it is 0.3 h faster than ALEXNet in training speed. It is verified that RSRM effectively improves the network training speed and accuracy of road image recognition.

1. Introduction

As car ownership has risen continuously, traffic jams, delays, and accidents spiraled upward. According to statistics [1, 2], 16.12% of traffic accidents on highways are attributed to slippery road conditions and the driver's response to changes in terrain caused by road damage. To improve vehicle safety, research on vehicle safety control has gradually changed focus from passive safety to active safety. As an important part of the vehicle's perception of the surrounding environment, road surface type recognition plays an important role not only in the power, smoothness, and comfort of intelligent driving vehicles, but also in vehicle safety.

In the 1960s, Wiesel and Hubel [3] found that their unique network structure could effectively reduce the complexity of the feedback neural network when they

studied the neurons used for local sensitivity and direction selection in the cortex of cats and proposed a convolutional neural network (CNN). Lecun et al. [4] made a great breakthrough in optical character recognition and computer vision by using a CNN, which promoted the development of computer vision. In recent years, CNNs have been widely used in many fields and have shown excellent performance in image target detection [5–7] and classification [8, 9]. The appearance of a CNN also provides a new solution for road condition recognition.

Several researchers and institutions have focused on pavement type identification and adhesion coefficient prediction. Chen [10] extracted the feature parameters of the gray-level cooccurrence texture matrix of the pavement image, studied the selection of pavement texture features, and achieved certain results. However, this method has the disadvantages of fewer image features and lower recognition

accuracy. Bekhtike and Kobayashi [11] used a camera to collect pavement images and evaluated the texture attributes obtained from the fractal dimension using Gaussian process regression for function approximation and predicted road types by fusing road texture features and vibration data received from motion. This method still has some limitations. For instance, when the background lighting changes obviously, motion blur occurs, or if the road is covered by rain, snow, or ice, it is difficult to accurately identify the road type.

Ward and Iagnemma [12] successfully classified asphalt, paved, and gravel roads with acceleration sensors. This method has drawbacks when the road surface roughness is similar, and it is obviously insufficient to use acceleration data to distinguish the road type. Alonso et al. [13] proposed a real-time acoustic pavement state recognition system based on tire noise, using a noise measurement system and a signal processing algorithm to classify the pavement state, and achieved accurate classification of wet and dry pavement states.

Neupane and Gharaibeh [14] proposed a method for detecting pavement types based on heuristic lidar and identified the pavement type by the mean and variance of the laser reflection intensity. This method is mainly used for asphalt pavement. Jonsson et al. [15] proposed road classification based on near infrared camera image spectral analysis, using KNN and support vector machine (SVM) methods to classify dry, wet, icy, and snowy roads and achieved certain results. Bystrov et al. [16] used automotive ultrasonic sensors to analyze reflected ultrasonic signals for road classification, with a recognition accuracy of up to 89%.

Meng [17] proposed a method based on the basic principles of machine learning to classify pavement types by combining data from vertical acceleration sensor signals and camera features. The accuracy of using an acceleration sensor or image data to identify road type was only 62% and 88%, respectively. When the two were combined, because of the small sample size, accuracy reached only 90%. Wang [18] classified and discriminated road images based on high-dimensional features and RBF neural networks and performed recognition experiments on eight different road images with an accuracy of approximately 78.4%. Based on the SVM, Zhao et al. [19] obtained the best classification model by PSO parameter optimization, classified the road types, and improved the recognition accuracy of the test image, achieving an accuracy rate of over 90% for the five basic road types.

Casselgren et al. [20] studied the light performance of asphalt pavements covered by water, ice, or snow. They conducted a detailed study on the changes in light intensity with the angle of incidence and spectrum changes and proposed two different wavebands to classify road conditions. Linton and Fu [21] described a networked vehicle-based winter road condition (RSC) monitoring solution that combines vehicle-based image data with data from road weather information systems. Jokela et al. [22] presented a method and evaluation results to monitor and detect road conditions (ice, water, snow, and dry asphalt).

The developed device is based on light polarization changes when reflected from the road surface. The recognition capability has been improved with texture analysis, which estimates the contrast content of an image, but the results show that the proposed solution does not currently adapt to different conditions perfectly well. Yeong [23] and Yu and Salari [24] developed a pothole detection system and method using 2D LiDAR. Caltagirone [25] developed a method for road detection in point cloud top-view images using fully CNN. However, according to the material presented in [26, 27], even LiDAR, which is the safest laser, can cause damage to the human eye during longer exposure (e.g., cataracts and burn of the retina). In the future, with the popularity of smart cars, this type of laser may be a problem. We consider a method to improve road condition recognition through image vision.

To summarize, most road recognition algorithms are based on traditional machine learning. Traditional machine learning extracts artificial image features as algorithm input. It was found that the process had a certain randomness, and the whole process including the classification algorithm was complex. To solve these problems, this paper proposes a road surface condition identification method based on an improved ALEXNet model, namely, the road surface recognition model (RSRM).

Therefore, the main contributions of this paper can be summarized as follows:

- (1) The ALEXNet [28] network model is pretrained on the ImageNet [29] dataset offline. The weights of the shallow network structure after training, including the convolutional layer, are saved and migrated to the proposed model. In addition, the fully connected layer fixed to the shallow network is replaced by 2 to 3, which improves the training accuracy and shortens the training time.
- (2) The traditional machine learning and improved ALEXNet model are compared, focusing on adaptability, prediction output, and error performance, among others.

2. Research Method for Identifying Road Surface Conditions Based on Improved ALEXNet Model (RSRM)

The traditional road type identification method has some limitations, such as a complex extraction process, weak adaptability, poor light robustness, low recognition accuracy, and difficulty in practical application. Meanwhile, the rapid development of artificial neural networks has also given birth to the progress of deep learning [30] in recent years. Common deep learning networks include autoencoders [31], deep belief networks [32], and CNNs. In deep learning, CNNs play a key role in image recognition. Road condition recognition belongs to the field of image recognition; therefore, in this study, the road image recognition model is built by combining CNNs and deep learning theory. With the help of CNN's self-learning and training of road image features, the actual road types can be identified.

2.1. Convolutional Neural Networks (CNN). CNN [33] has high efficiency and accuracy in image recognition, which is due to the shared parameters of convolutional kernels in the hidden layer and sparsity of interlayer connections. A CNN model is generally formed by alternately stacking convolutional layers and pooling layers, and the specific operation for input data is saved in the weight of this layer. The loss function is used to evaluate the difference between the output and target values. The optimizer uses the difference between the target value and the output value as the feedback signal to update the weight value through the back-propagation algorithm [34] and finally reduces the loss value corresponding to the current target, which makes the network prediction more accurate. The feature values of the last layer of the pooling layer generate a list of vectors through the fully connected layer and input them to SoftMax [35], for classification and recognition. The CNN training process is shown in Figure 1.

2.1.1. Convolutional Layers. The convolutional layers principally perform convolution operation on the image or feature map, which is input into the convolution layer, to extract feature and output the convoluted feature map. Therefore, as shown in equation (1), each feature map of convolution layer is obtained by combining and calculating multiple feature maps output from the previous layer:

$$X_n^l = f \left(\sum_{i \in M_n} X_i^{l-1} * K_{in}^l + b_n^l \right), \quad (1)$$

where M_n is the feature map set filtered from the input feature map, X_n^l is the n th feature map in the l th layer, K_{in}^l is the i th element of the n th convolution kernel in the l th layer, b_n^l is the n th offset of the l th layer, and “*” is the process of convolution.

2.1.2. Pooling Layers. The pooling layer, also known as the lower sampling layer, is mainly used to reduce the calculation amount of feature extraction. The pooling layer retains the number of feature maps but changes the size of the feature maps; equation (2) represents the calculation process of the sampling layers.

$$X_n^l = f(\beta_n^l \text{down}(X_n^{l-1}) + b_n^l), \quad (2)$$

where $\text{down}(\cdot)$ is the lower sampling (pooling) function, β_n^l is the n th multiplication offset of the l th layer, and b_n^l is the n th offset of the l th layer.

The lower sampling function is largely divided into mean-pooling and max-pooling. Mean-pooling is to calculate the average of all elements in the pooling area.

$$P_n = \frac{1}{|R_n|} \sum_{i \in R_n} c_n. \quad (3)$$

The max-pooling is to select the maximum element in the pool area.

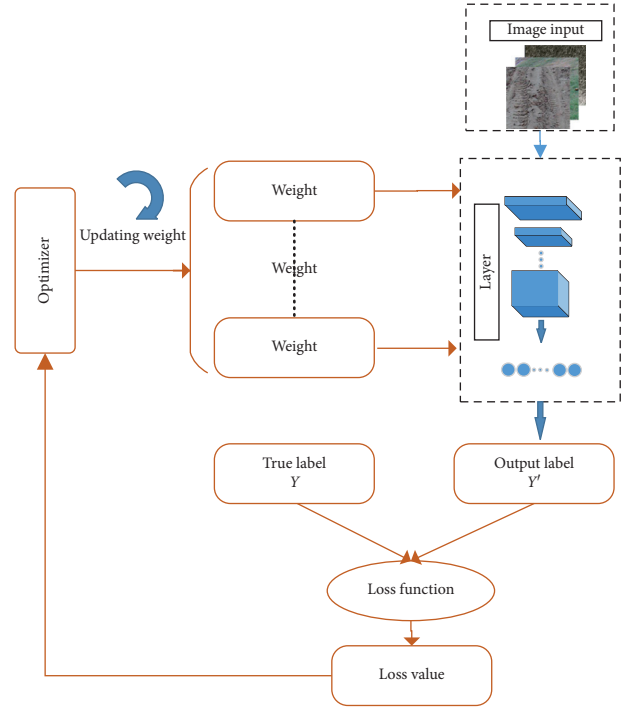


FIGURE 1: The training process of road surface image based on deep learning.

$$P_n = \max_{i \in R_n} c_i, \quad (4)$$

where R_n is the n th pooling area in the feature map and c_i is the i th pixel value in R_n .

2.1.3. Fully Connected Layers. The fully connected layers generally locate at the last part of the hidden layers in CNN. The fully connected layers form a multilayer perceptron like the shallow neural network, which nonlinearly combines the feature vectors output by the convolutional layer and the pooling layer to get the output.

2.1.4. Output Layers. The output layers in CNN are usually behind the fully connected layers. For image classification problems, the output layers use a logical function or a normalized exponential function (SoftMax function) to output classification labels. The range of the multi-classification label y in SoftMax regression is $y \geq 2$. The training sample set is composed of k labeled samples:

$$T = \{(x_{(1)}, y_{(1)}), (x_{(2)}, y_{(2)}), \dots, (x_{(k)}, y_{(k)})\}, \quad (5)$$

where $y_{(j)} \in \{1, 2, \dots, k\}$ is the classification labels, and $x_{(j)}$ is the sample set. j represents different classifications, and it is estimated probability value. The probability that a single sample is classified into class K is

$$P(y = j|x), \quad (j = 1, 2, \dots, k). \quad (6)$$

The regression sample set is transformed into a k -dimensional probability vector, and it is given by

$$h(x_i|\theta) = \begin{bmatrix} p(y_{(i)} = 1|x_{(i)}, \theta) \\ p(y_{(i)} = 2|x_{(i)}, \theta) \\ \vdots \\ p(y_{(i)} = k|x_{(i)}, \theta) \end{bmatrix} = \frac{1}{\sum_{j=1}^k e^{j^T x_{(i)}}} \begin{bmatrix} e^{\theta_1^T x_{(i)}} \\ e^{\theta_2^T x_{(i)}} \\ \vdots \\ e^{\theta_k^T x_{(i)}} \end{bmatrix}, \quad (7)$$

where $\theta = [\theta_1^T, \theta_2^T, \dots, \theta_k^T]$ is learning parameters, and $\theta_1^T, \theta_2^T, \dots, \theta_k^T \in \mathbf{R}^{n+1}$; the purpose of $(1/\sum_{j=1}^k e^{\theta_j^T x_{(i)}})$ is to normalize the probabilities and make the sum of the probabilities be 1.

Through the training of sample set, the optimizer adjusts parameters to minimize model loss function value, and its loss function formula is defined as

$$J(\theta) = -\frac{1}{m} \left[\sum_{i=1}^m \sum_{j=1}^k 1\{y_{(i)} = j\} \log \frac{e^{\theta_j^T x_{(i)}}}{\sum_{i=1}^k e^{\theta_i^T x_{(i)}}} \right]. \quad (8)$$

2.2. ALEXNet Network. The typical CNN models usually include GoogLeNet [36], VGG-16 [37], and ALEXNet. GoogLeNet and VGG-16 have 22 and 16 layers, respectively. Theoretically, the deeper the number of model layers, the better the classification effect.

As a classic model, ALEXNet accelerates the development of deep learning, which is a milestone in image recognition. Before the research, we have done the comparison between ALEXNet and VGG, GoogleNet, and other networks, and ALEXNet network can reach a higher recognition accuracy in a shorter time.

Second, theoretically, the deeper the model layer is, the better the classification effect is. However, the training process of deep convolution network is extremely difficult. For example, many parameters lead to the disappearance of backpropagation gradient and overfitting. At the same time, the deeper network often needs to consume more computing resources. ALEXNet can meet the accuracy of road image recognition, while reducing computer resources. So, the diversity of road images is low, and ALEXNet can achieve higher recognition accuracy and occupy less computer resources.

Third, the road image is relatively simple, and the latest network is usually to solve more complex image classification problems. ALEXNet has been able to solve the problem of road condition image recognition extremely well.

This is due to several advantages of the ALEXNet network:

- (1) In the training process, dropout is used to randomly ignore some neurons to avoid overfitting the model.
- (2) Samples are data augmented [38] to expand the samples with insufficient training images.
- (3) Rectified Linear Units (ReLU) [39] are used as the excitation function of the network, which improves its nonlinearity and solves the problem of gradient

dispersion. To solve the problem of gradient dispersion, ALEXNet adopts the ReLU activation function. ReLU is defined as follows:

$$\text{ReLU}(x) = \max(x, 0). \quad (9)$$

In Figure 2, comparing ReLU and sigmoid [40] activation function curves, it shows that when x is greater than 0, the ReLU gradient value is always a constant of 1. The derivative of the sigmoid function is like the curve shape of the Gaussian function, but not constant. The derivative at both ends of the sigmoid curve becomes smaller. Therefore, the network with ReLU as an activation function converges quickly, which is helpful in accelerating training.

Figure 3 and Table 1 show the structure and parameters of ALEXNet. The model is mainly composed of five convolutional layers and three fully connected layers. The number of convolution kernels in five convolution layers is 96, 256, 384, 384, and 256, respectively. The role of the pooling layer is mainly to reduce the size of the feature image after convolution. The nodes of the three fully connected layers are 4096, 4096, and 1000, respectively. SoftMax can classify 1000 categories.

2.3. Road Surface Recognition Model Based on RSRM. The ALEXNet network was pretrained on the ImageNet database with at least one million images offline, and the weights and parameters of each layer were obtained after training. The trained network has a strong ability to learn features, especially curves, edges, and contours of an image. To improve the efficiency of network training and reduce the training time, this study takes the trained ALEXNet network as the pretrained model and transfers its parameters to the RSRM using fine-tuning transfer learning [41]. ALEXNet, SVM, and BP use the classic structure. SVM algorithm is based on the characteristics of the road image for road color and texture feature extraction experiments.

Similarly, RSRM consists of a convolutional layer, pooling layer, fully connected layer, and SoftMax classification layer. By analyzing the characteristics of actual pavement images, nine typical pavement types are selected, as shown in Figure 4, focusing on nine typical road surface types; therefore, 9-label SoftMax is used to replace the original classifier in the ALEXNet network. In addition, as shown in Figure 5, two fully connected layers are trained on the actual road pavement test set and to replace the original three fully connected layers. The number of nodes in the two fully connected layers are 4096 and 1000, respectively.

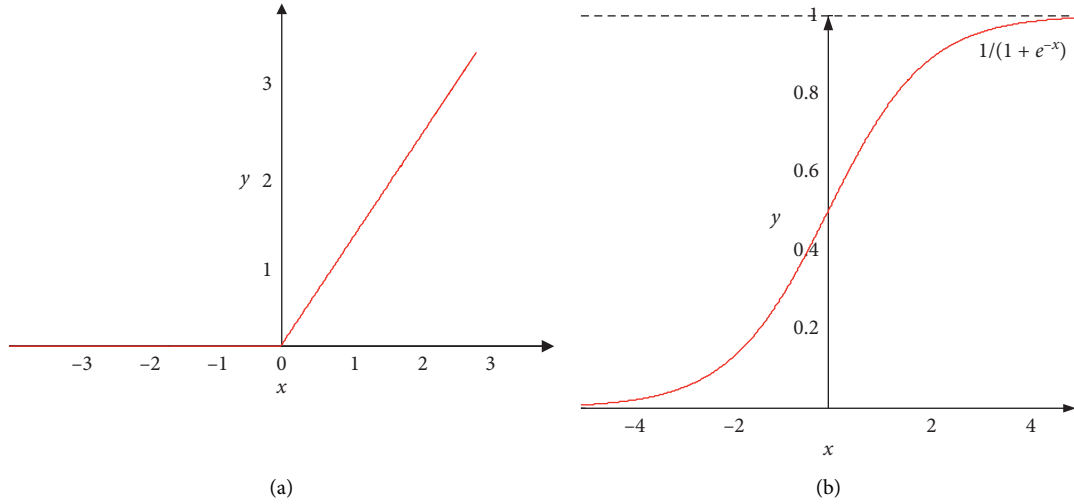


FIGURE 2: Activation function. (a) ReLU function image. (b) Sigmoid function image.

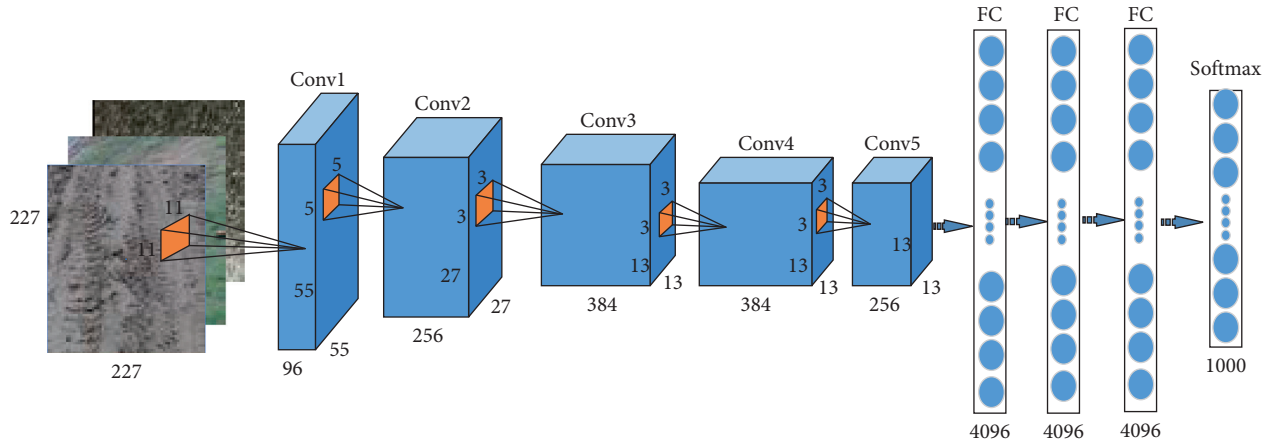


FIGURE 3: The structure of AlexNet network models.

TABLE 1: AlexNet structure parameters.

Name	Parameter	Stride	Number of convolution kernels
Conv 1	$11 \times 11 \times 3$	4	96
Pool 1	$1 \times 3 \times 3 \times 1$	2	—
Conv 2	$5 \times 5 \times 48$	1	256
Pool 2	$1 \times 3 \times 3 \times 1$	2	—
Conv3	$3 \times 3 \times 256$	1	384
Conv4	$3 \times 64 \times 384$	1	384
Conv5	$3 \times 3 \times 384$	1	256
Pool5	$1 \times 3 \times 3 \times 1$	2	—
Fc6	4096	—	—
ReLU6	ReLU	—	—
Dropout	0.5	—	—
Fc7	4096	—	—
ReLU7	ReLU	—	—
Dropout	0.5	—	—
Fc8	1000	—	—
Prob	SoftMax	—	—
Output	1000	—	—

Through the above steps, the problem of road surface image classification and recognition is solved.

3. Experimental Settings

3.1. Road Surface Data Acquisition System. The road collection test vehicle was a sedan with a length of 3564 mm, width of 1620 mm, and height of 1527 mm. Its wheelbase was 2340 mm. The camera model was LeTMC-520. As shown in Figure 6, the camera was installed at the air intake grille in the front of the vehicle, at an angle of -10° from the horizontal grille. The installation height from the ground was 350 mm. In this study, considering the complex weather conditions in the actual driving process, three typical weather conditions, namely, cloudy, sunny, and rainy, were selected for road image data collection. Note that the images of the actual road test set are all taken by the vehicle during driving.

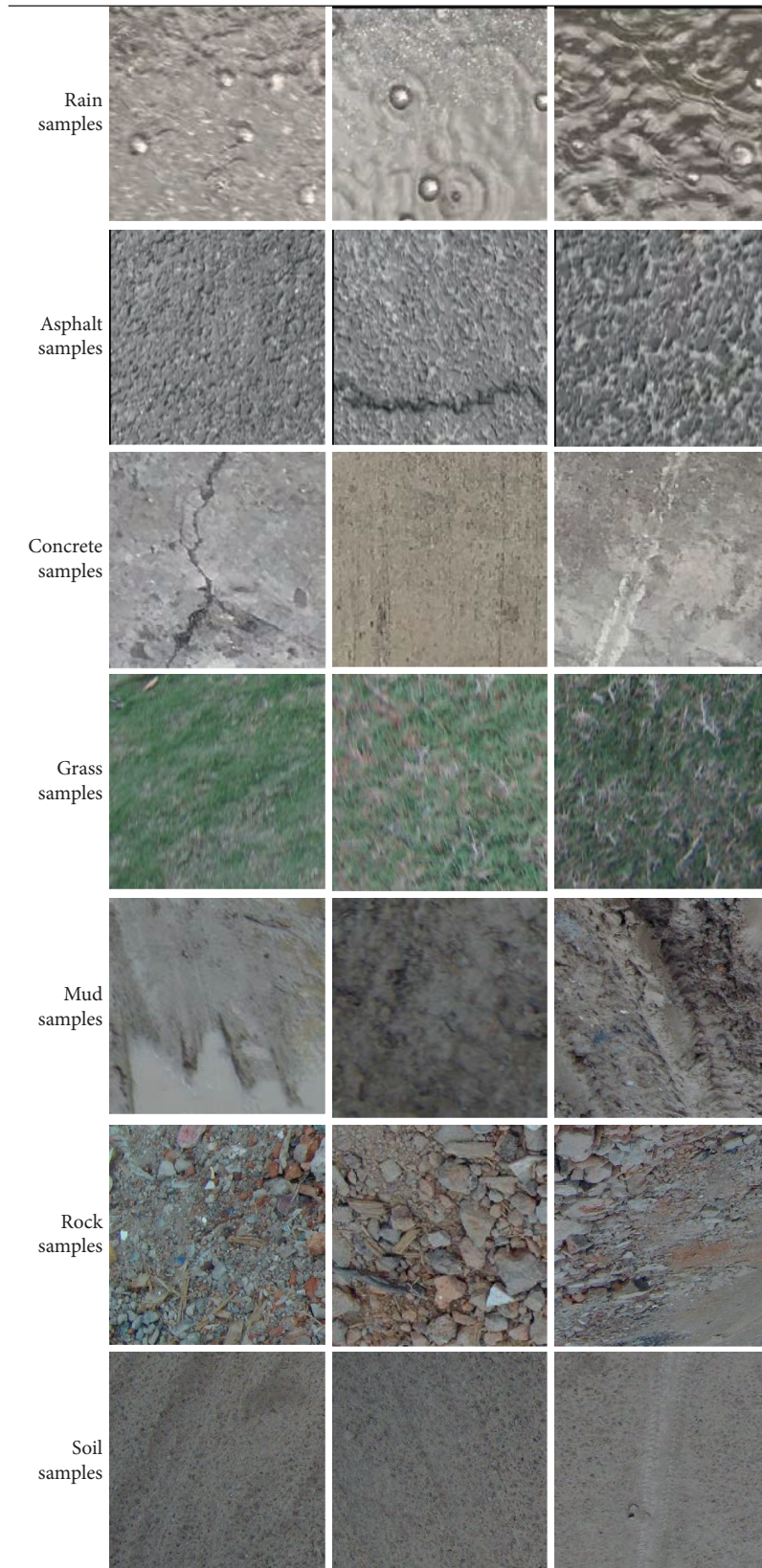


FIGURE 4: Image samples of road surface state.

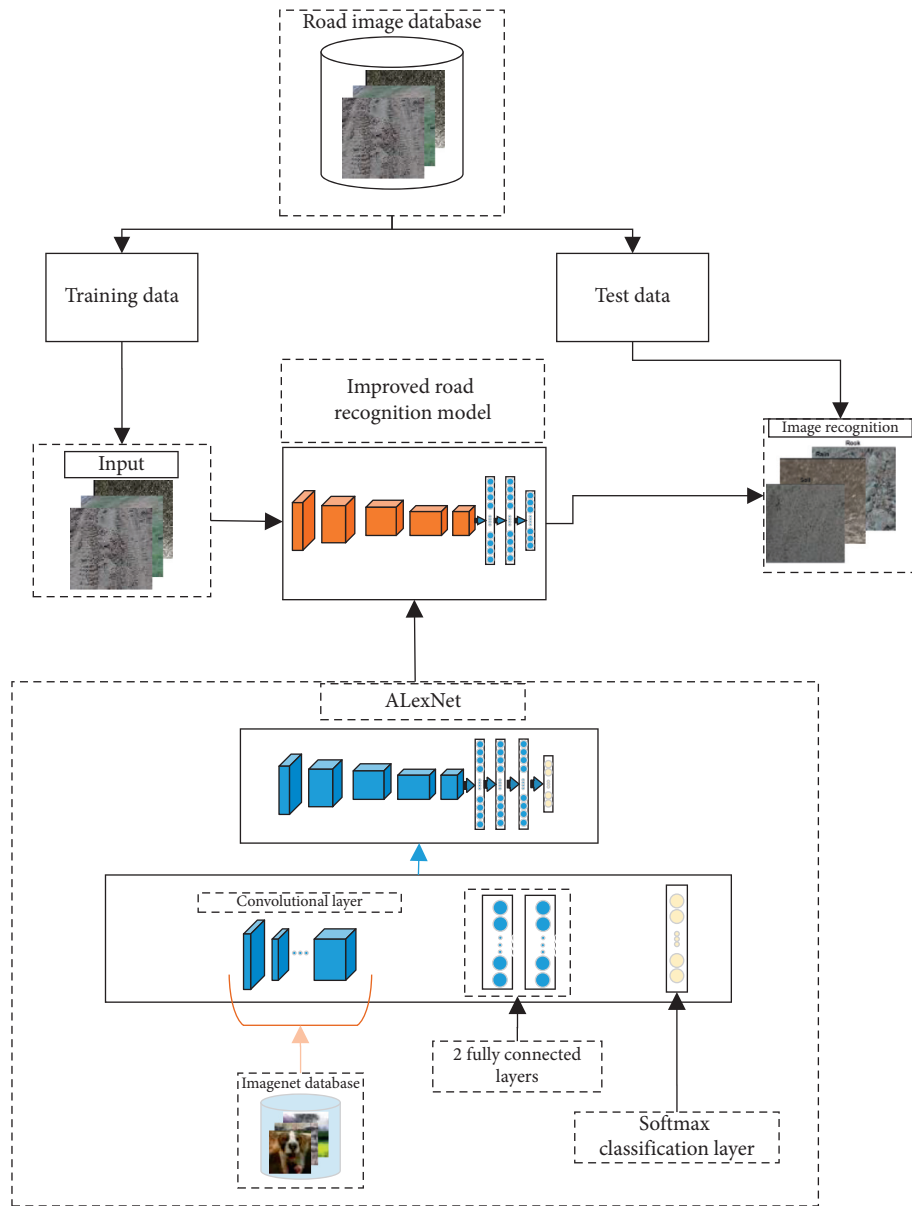


FIGURE 5: Structure of the road surface recognition model.

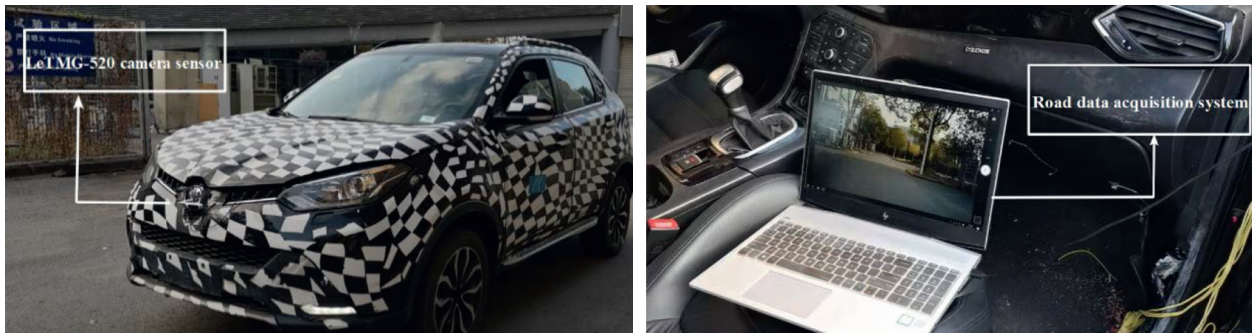


FIGURE 6: Camera installation and data acquisition.

In addition, a road surface data analysis system server configuration was performed on a desktop computer with a 64 bit operating system, 16 GB of memory, an AMD Ryzen 5 3600 6-Core Processor, and a GeForce GTX 1660 graphics processing unit.

3.2. Establishment of Road Surface Image Database. The image standards were selected according to the typical pavement types: asphalt, concrete, grass, mud, rain, rock, soil, wet asphalt, and wet concrete, and the images with clear quality were used for the road surface image database (RSID). The sample size of each pavement was 2000, in which the training set and test set were divided in a ratio of 7:3.

3.3. Experimental Procedure

Step 1: Image preprocessing.

Scaling and cropping operations are performed on all road surface images to ensure a uniform image size, which can meet the requirements of the neural network module in MATLAB.

Step 2: Building the training set and test set.

RSID is divided into the training set and test set in a ratio of 7:3.

Step 3: Building the RSRM.

Focusing on nine typical road surface types, the 9-label SoftMax is used to replace the original classifier in the ALEXNet network. The next step is to use the trained ALEXNet network as the pretrained model and transfer its parameters to the RSRM using fine-tuning transfer learning.

Step 4: Model training

Model training that uses the stochastic approach initializes the model parameters; sets the momentum parameters, learning rate, and training time; and freezes the parameters of the five convolutional layers and pooling layers. Through the above, we replace the parameters of the two fully connected layers and 9-label SoftMax with a fresh new one.

Step 5: Model testing.

The remaining 30% of the RSID was used as a test set to verify the accuracy and speed of the RSRM.

4. Results and Analysis of Experiment

4.1. Experiment of Road Image Feature Extraction. The role of the convolution layer is to extract features by performing a convolution operation on the image or feature map. First, we pretrained the improved ALEXNet network model (5 convolutional layers) on the ImageNet dataset. Second, the weights of the shallow network structure after training were saved and transferred to the RSRM. Finally, to observe the feature extraction effect of RSRM more clearly, taking the mud image as an example, the output features of each



FIGURE 7: Image of mud pavement after preprocessing.

convolution layer were visualized. Figure 7 shows the mud pavement image after preprocessing.

As shown in Figure 8(a), the preprocessed mud image is extracted with 96 feature maps through Conv1. The convolution layer mainly extracts edges and details of the image. After several convolution kernel operations in the convolution layer, the image retains most of the information of the original image. As shown in Figure 8(b), the convolved image is processed by the ReLU1 activation function, and the edge information and detailed information of the mud surface road image are more obvious. Figures 8(c) and 8(d) show the feature map after Conv3 and relu3. It can be seen from the figure that the convolution kernel can extract more edge information, and the outline of the mud road surface image is clearer. Figures 8(e) and 8(f) show the feature map after Conv5 and relu5. It also reveals that as the number of convolutional layers increases from the first layer to the fifth layer, the resolution ratio of the image decreases, and the image output from the convolutional kernel becomes increasingly abstract.

According to the above image feature extraction experiments, the convolution layer integrates shallow features or underlying features to form more abstract features. This makes the expression of road information more comprehensive and can also use high-level abstract features for pavement classification and recognition.

4.2. Experiments of Road Surface Type Recognition Based on RSRM. RSID contains 18000 images of nine pavement types, such as asphalt, wet asphalt, rain, concrete, wet concrete, soil, mud, grass, and rock. To verify the validity of the RSRM proposed in this paper, 70% of RSID were randomly selected as the training set, with a total of 12600 pieces, and the remaining 30% was used as the test set. There were 600 images for each type of pavement in the test set, for a total of 5400 pavement images.

Table 2 shows the RSRM training parameter setting. The test tolerance is the number of iterations that the loss of test set before network training stops can be greater than or equal to the previously smallest loss. This can stop training by setting the test tolerance when test loss is no longer

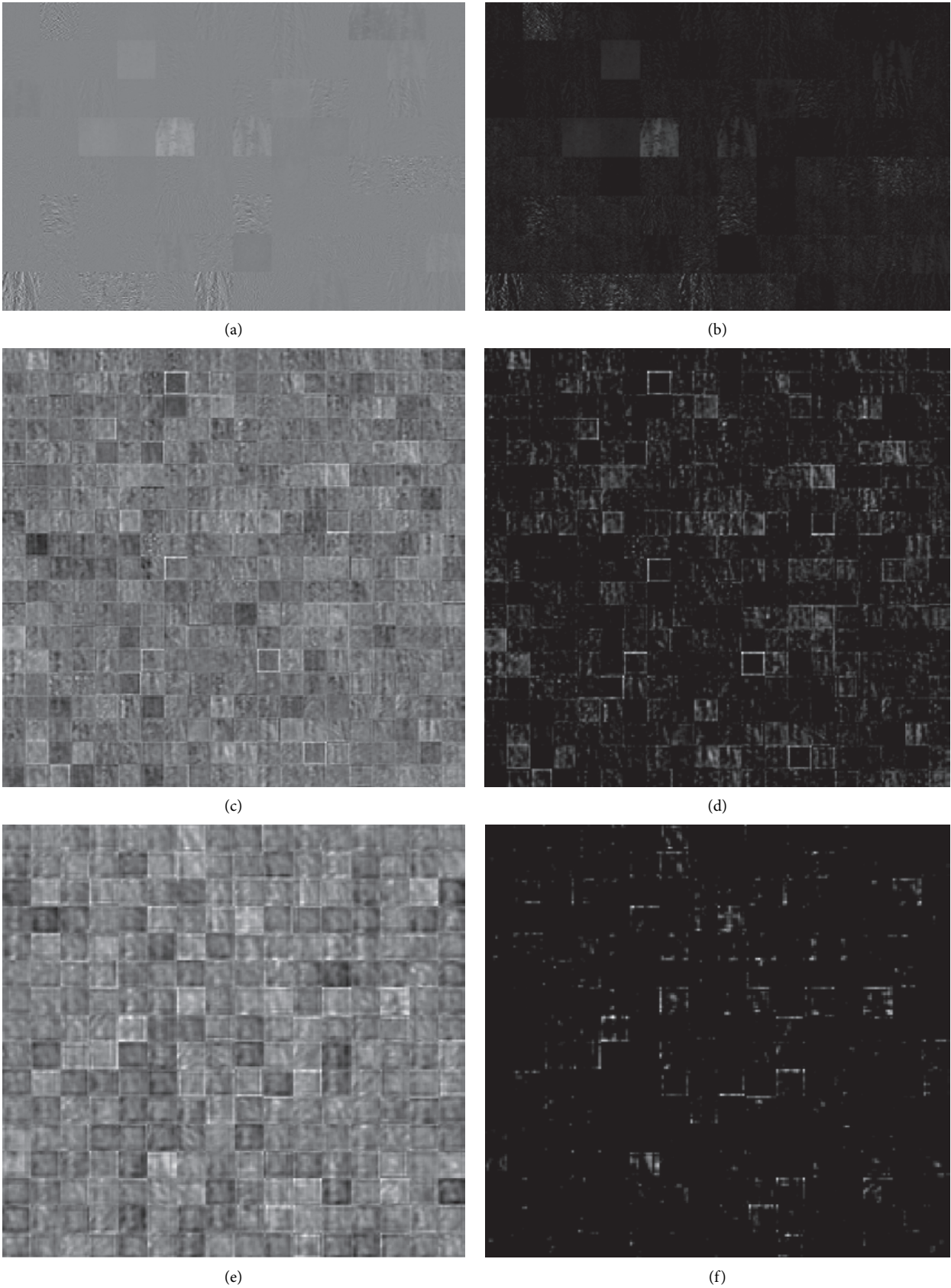


FIGURE 8: Feature extraction results of different Conv and ReLu for mud pavement images. (a) Conv1. (b) ReLu1. (c) Conv3. (d) ReLu3. (e) Conv5. (f) ReLu5.

TABLE 2: Training parameter setting.

Verification frequency	Minibatch size	Epoch	Learning rate	Test tolerance
20/iteration	128/images	25	0.001	5/iteration

reduced, to avoid overfitting, save computer memory and improve training speed.

Table 3 shows the classification results of test samples based on RSRM in this research. Table 4 lists the recognition results of some image based on RSRM.

According to Step 2, there are 600 samples per category in the test set; this is true for the asphalt pavement type. As can be seen in Table 3, an asphalt pavement image (600 samples) was misidentified, and the recognition accuracy was 99.8%. This is because part of the asphalt pavement presents a dry-wet state, which makes it extremely similar to the image characteristics of asphalt pavement; thus, it is misidentified as wet asphalt. A total of 598 concrete pavement samples were correctly identified, and the remaining two were identified as soil and mud pavements, with an accuracy rate of 99.7%. The reason is that the color and image texture of some concrete, soil, and mud pavements are similar under dry conditions. The number of rain pavement samples correctly identified is 598, with the remaining two misidentified as wet asphalt and wet concrete pavements; meanwhile, the identification accuracy rate was 99.7%. For soil pavement, 599 samples were correctly identified, and the remaining one was classified as mud. A wet soil road surface often forms the mud surface, and the high probability of these two pavement features cooccurring in a single image is the main factor leading to false positives. The total number of wet concrete surfaces is 600, of which 580 are correctly recognized, 9 are identified as soil, one is identified as rock pavement, and the last 10 are identified as mud pavement. Thus, the recognition accuracy of wet concrete pavement is 96.7%. This is because the color of the wet concrete pavement is brown-gray after being wet. The recognition accuracy of grass, rock, and wet asphalt are higher than other surfaces, which is due to the significant difference in color and texture features compared to other road images.

4.3. Experiments of Classification Method Comparison. In this study, RSRM is compared against the ALEXNet model, support vector machines (SVM), and backpropagation (BP) neural networks. The results are shown in Table 5. According to previous research, color and texture are the main features of road images. The SVM [42] classification model needs to extract road image features manually. In the three-color spaces of HSV, RGB, and YCM, there are nine color components of the road image, namely H, S, V, R, G, B, Y, C, and M. The gray-level cooccurrence matrix is used to extract four texture similar information of road surface images, such as contrast, correlation energy, and entropy. The BP neural network [43–45] has five layers, the number of nodes in each layer is 100, and the optimization algorithm uses stochastic gradient descent.

In this section, RSRM is compared with the BP neural network, SVM, and ALEXNet models, focusing on the

TABLE 3: The classification results of test samples based on RSRM.

		Confusion matrix										
		Asphalt	Concrete	Grass	Mud	Rain	Rock	Soil	Wet asphalt	Wet concrete		
Output class	Asphalt	599 11.1%	0 0.0%	0 0.0%	0 0.0%	1 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	99.8% 0.2%
	Concrete	0 0.0%	598 11.1%	0 0.0%	0 0.0%	1 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	99.8% 0.2%
	Grass	0 0.0%	0 0.0%	600 11.1%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
	Mud	0 0.0%	1 0.0%	0 0.0%	598 11.1%	0 0.0%	0 0.0%	1 0.0%	0 0.0%	10 0.2%	98.0% 2.0%	
	Rain	0 0.0%	0 0.0%	0 0.0%	0 0.0%	598 11.1%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%	
	Rock	0 0.0%	0 0.0%	0 0.0%	1 0.0%	0 0.0%	600 11.1%	0 0.0%	0 0.0%	1 0.0%	99.7% 0.3%	
	Soil	0 0.0%	1 0.0%	0 0.0%	1 0.0%	0 0.0%	0 0.0%	599 11.1%	0 0.0%	9 0.2%	98.2% 1.8%	
	Wet asphalt	1 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	600 11.1%	0 0.0%	99.8% 0.2%	
	Wet concrete	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	580 10.7%	100% 0.0%	
			99.8% 0.2%	99.7% 0.3%	100% 0.0%	99.7% 0.3%	99.7% 0.3%	100% 0.0%	99.8% 0.2%	100% 3.3%	96.7% 0.5%	99.5% 0.5%
		Asphalt	Concrete	Grass	Mud	Rain	Rock	Soil	Wet asphalt	Wet concrete		
		Target class										

analysis of model prediction output, error performance, training time, and detection time.

As shown in Figures 9 and 10, RSRM significantly improves accuracy of road surface identification compared to ALEXNet. Specifically, RSRM converged at 216 iterations, realizing an accuracy of 96.38%. However, the ALEXNet network has yet to converge after 500 iterations. Therefore, transfer learning and optimization of the fully connected layer can effectively improve the training efficiency and accuracy of the model. The ALEXNet model requires a longer training time and larger dataset to match the accuracy of RSRM.

Figure 11 and Table 5 illustrate the identification accuracy of different methods. The average recognition accuracy of BP, SVM, ALEXNet, and RSRM was 92.84%, 89.59%, 97.57%, and 99.48%, respectively. The accuracy of RSRM and ALEXNet is more than 95%, which shows the superiority of deep learning methods. Traditional machine learning methods, such as SVM and BP neural networks, are not suitable for representing variations in illumination intensity due to their manual features. The SVM classifier is suitable for small datasets, which is why it has not achieved good results in road datasets. Table 6 shows the average time taken by each learning model to classify a test image. The test times of all models for a given road image are almost the same. The results show that the accuracy of the deep learning model is higher than that of the traditional machine learning method.

TABLE 4: Results of the test sample classification.

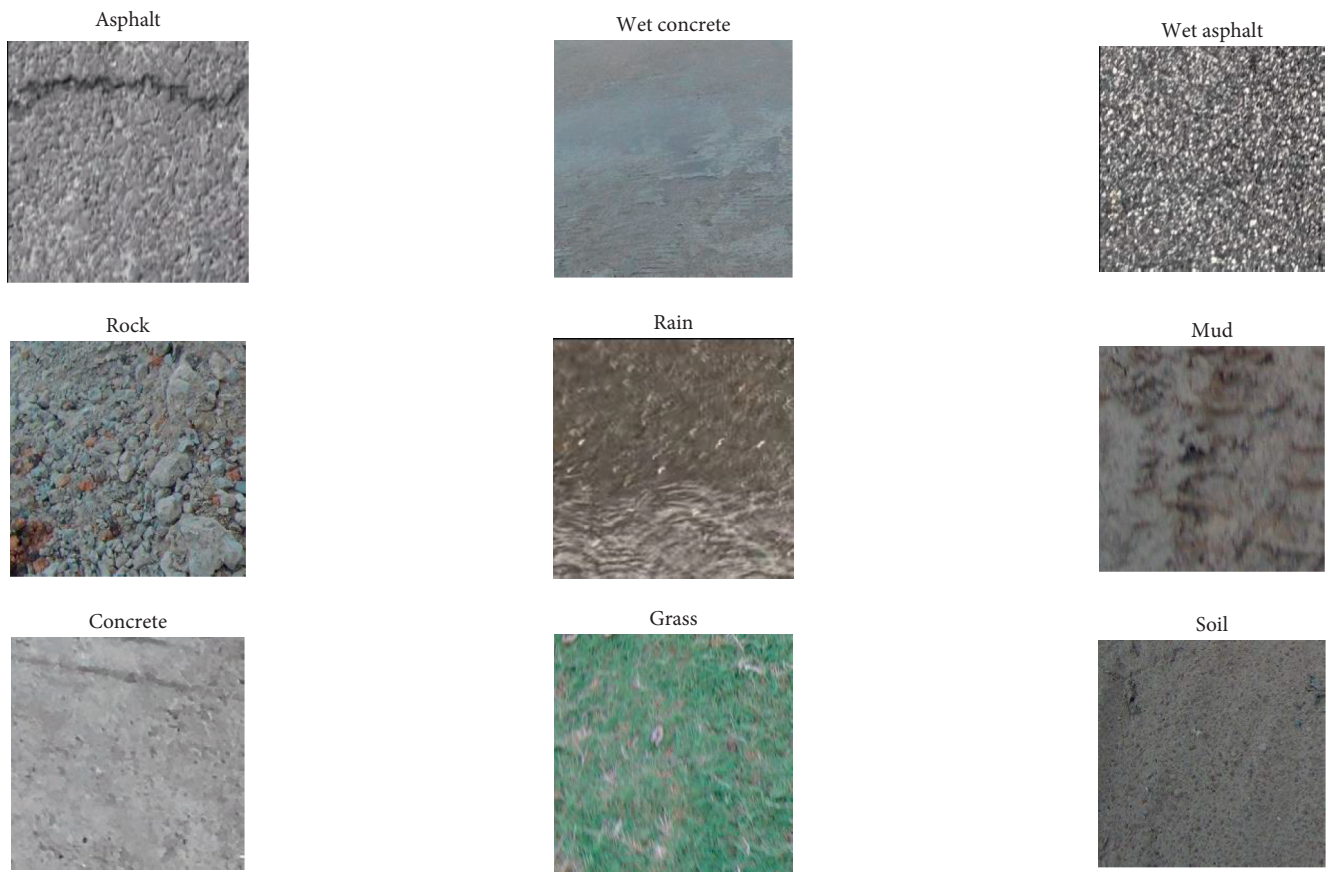


TABLE 5: The accuracy of different model identification.

Model (%)	SVM	BP	ALexNet	RSRM
Training accuracy	91.72	93.36	98.36	100
Test accuracy	89.59	92.84	97.57	99.48

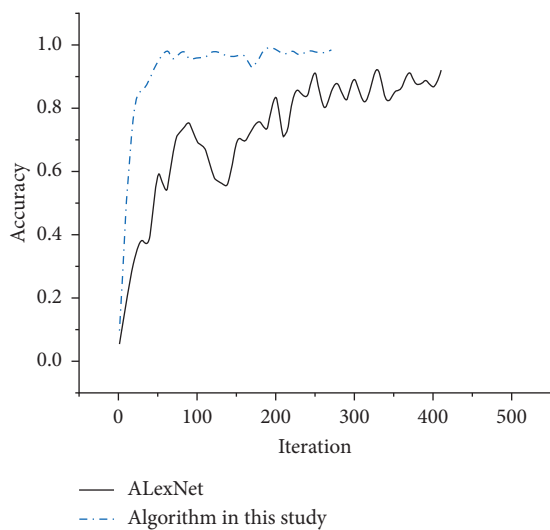


FIGURE 9: Accuracy during validation.

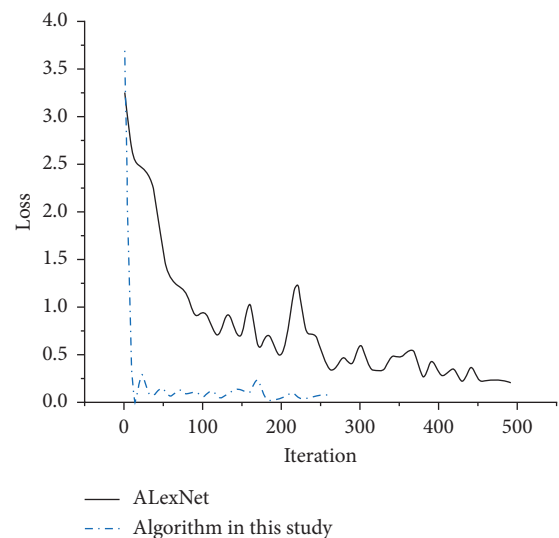


FIGURE 10: Loss during validation.

SVM is effective in dealing with small-scale datasets, which is difficult to implement for large-scale training samples. In the process of image classification based on BP, the upper layer of neurons and the next layer of neurons are fully connected, which leads to excessive

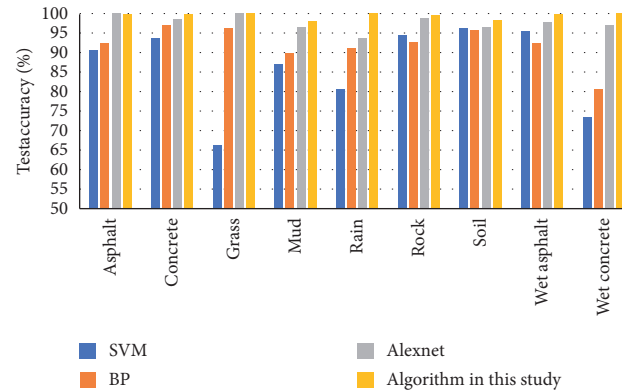


FIGURE 11: Test results of different pavement identification models.

TABLE 6: Training time and individual image recognition time for different models.

Model	SVM	BP	ALexNet	RSRM
Training time/h	1.6	2.2	1	0.4
Recognition time/s	0.15	0.16	0.15	0.14

training weight and overfitting. However, a CNN can effectively reduce the training weight and improve the training speed using a convolution operation.

In addition, this study proposes a method for testing tolerance thresholds to stop model training and reduce the number of fully connected layers. Based on this, the SoftMax classifier for nine labels is designed. The training time for the proposed method was 1.6 h, BP was 2.2 h, ALexNet was 1 h, and RSRM required 0.4 h (RSRM does not include pre-training time), and it took 0.14 s to classify a test image. The training speed of RSRM is four times that of SVM and five times that of BP. Meanwhile, the recognition accuracy was 1.91% higher than that of ALexNet, 6.64% higher than BP, and 9.89% higher than SVM. RSRM can effectively improve the training efficiency and accuracy of the model.

In summary, the BP neural network is not suitable for recognizing multiple ranges of road image databases because of the large number of neurons, the number of network layers cannot be too large, and the computing time is long, which can easily lead to overfitting and inconvenience in processing high-dimensional data. SVM feature extraction is complex and only suitable for small datasets. The proposed method not only achieves fast and high-precision recognition of road surface types in a short training time but also meets the perception requirements of actual road conditions.

5. Conclusion

This paper presents a pavement identification method based on an improved ALexNet model. First, the ALexNet network model is pretrained on the ImageNet dataset offline. Second, the weights of the shallow network structure after training, including the convolutional layer, are saved, and migrated to the proposed model. In addition, the fully connected layer fixed to the shallow network is replaced by 2 to 3, which

improves the training accuracy and shortens the training time, and the 9-label SoftMax replaces the original classifier in the ALexNet network. In addition, the proposed method is compared with the BP neural network, SVM, and ALexNet models, focusing on the prediction output, error performance, and rapidity of the model. The results show that the recognition accuracy of RSRM is 99.48%, which is higher than that of ALexNet, BP, and SVM by 1.91%, 6.64%, and 9.89%, respectively. Moreover, this paper proposes a method for testing tolerance thresholds to stop model training and reduce the number of fully connected layers, which can save 0.6 h of training time and increase the training speed to four times that of SVM and five times that of BP. In conclusion, the deep learning model not only has higher accuracy than the traditional machine learning method but also can achieve higher recognition accuracy in a shorter time, which can meet the perception requirements of actual road conditions. The research method is not only suitable for road recognition, but also suitable for human-vehicle-road collaborative perception of the vehicle environment.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was funded by the National Natural Science Foundation of China, under grant no. 51575232, and the Shanghai Youth Science and Technology Talents Sailing Project, under grant no. 19YF1434600.

References

- [1] X. Sun, *Research on Risk Coupling of Highway Traffic Safety Based on Catastrophe Theory*, Beijing Jiaotong University, Beijing, China, 2015.
- [2] S. D. Gleave, R. Frisoni, and F. Dionori, "EU road surfaces: economic and safety impact of the lack of regular road

- maintenance, study, transport and tourism,” *European Union*, vol. 24, no. 135, pp. 897–906, 2014.
- [3] T. N. Wiesel and D. H. Hubel, “Effects of visual deprivation on morphology and physiology of cells in the cat’s lateral geniculate body,” *Journal of Neurophysiology*, vol. 26, no. 6, pp. 978–993, 1963.
 - [4] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
 - [5] R. Girshick, “Rich feature hierarchies for accurate object detection and semantic segmentation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580–587, Columbus, OH, USA, June 2014.
 - [6] W. Liu, D. Anguelov, D. Erhan, C. Szegedy et al., *SSD Single Shot Multibox Detector*, *European Conference on Computer Vision*, Springer, Cham, Berlin, Germany, 2016.
 - [7] J. Redmon, “You only look once: unified, real-time object detection,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779–788, Las Vegas, NV, USA, June 2016.
 - [8] D. Lan, “Target detection method based on convolutional neural network for SAR image,” *Journal of Electronics Information Technology*, vol. 38, no. 12, pp. 3018–3025, 2016.
 - [9] H. Zhang, L. I. Yumei, and F. U. Yingying, “Research on natural scene image classification based on haar-CNN model,” *Journal of Sichuan Normal University (Natural Science)*, vol. 40, no. 1, pp. 119–126, 2017.
 - [10] Y. Chen, “Image analysis applied to black ice detection,” *Applications of Artificial Intelligence IX. International Society for Optics and Photonics*, vol. 1468, pp. 551–562, 1991.
 - [11] M. A. Bekhti and Y. Kobayashi, “Prediction of vibrations as a measure of terrain traversability in outdoor structured and natural environments,” *Image and Video Technology*, vol. 34, no. 8, pp. 282–294, 2015.
 - [12] C. C. Ward and K. Iagnemma, “Speed-independent vibration-based terrain classification for passenger vehicles,” *Vehicle System Dynamics*, vol. 47, no. 9, pp. 1095–1113, 2009.
 - [13] J. Alonso, J. M. López, I. Pavón, M. Recuero et al., “On-board wet road surface identification using tyre/road noise and support vector machines,” *Applied Acoustics*, vol. 76, pp. 407–415, 2014.
 - [14] S. R. Neupane and N. G. Gharaibeh, “A heuristics-based method for obtaining road surface type information from mobile lidar for use in network-level infrastructure management,” *Measurement*, vol. 131, pp. 664–670, 2019.
 - [15] P. Jonsson, J. Casselgren, and B. Thörnberg, “Road surface status classification using spectral analysis of NIR camera images,” *IEEE Sensors Journal*, vol. 15, no. 3, pp. 1641–1656, 2014.
 - [16] A. Bystrov, E. Hoare, T. Y. Tran, N. Clarke, M. Gashinova, and M. Cherniakov, “Road surface classification using automotive ultrasonic sensor,” *Procedia Engineering*, vol. 168, pp. 19–22, 2016.
 - [17] W. Meng, “Research on vehicle pavement type recognition technology based on machine learning,” *Journal of Military Science and Technology*, vol. 38, no. 8, pp. 1642–1648, 2017.
 - [18] W. F. Wang, “Classification of slippery road images based on high-dimensional features and RBF neural network,” *Journal of Transport Information and Safety*, vol. 31, no. 2, pp. 32–35, 2013.
 - [19] J. Zhao, H. Wu, and L. Chen, “Advances in traffic safety methodologies and technologies,” *Journal of Advanced Transportation*, vol. 2017, Article ID 6458495, 21 pages, 2017.
 - [20] J. Casselgren, M. Sjö Dahl, and J. LeBlanc, “Angular spectral response from covered asphalt,” *Applied Optics*, vol. 46, no. 20, pp. 4277–4288, 2007.
 - [21] M. A. Linton and L. Fu, “Connected vehicle solution for winter road surface condition monitoring,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2551, no. 1, pp. 62–72, 2016.
 - [22] M. Jokela, M. Kutila, and L. Le, “Road condition monitoring system based on a stereo camera,” in *Proceedings of the IEEE 5th International Conference on Intelligent Computer Communication and Processing*, pp. 423–428, IEEE, Cluj-Napoca, Romania, August 2009.
 - [23] B. K. Yeong, “Pothole detection system using 2D LiDAR and camera,” in *Proceedings of the Ninth International Conference on Ubiquitous and Future Networks (ICUFN)*, pp. 744–746, IEEE, Milan, Italy, July 2017.
 - [24] X. Yu and E. Salari, “Pavement pothole detection and severity measurement using laser imaging,” in *Proceedings of the IEEE International Conference on Electro/Information Technology*, pp. 1–5, IEEE, Mankato, MN, USA, May 2011.
 - [25] L. Caltagirone, “Fast LIDAR-based road detection using fully convolutional neural networks, intelligent vehicles symposium,” in *Proceedings of the 2017 IEEE Intelligent Vehicles Symposium (iv)*, Los Angeles, CA, USA, June 2017.
 - [26] K. Schulmeister and G. Veas, “Principles for consistent application of the IEC laser product safety standard based on legal requirements,” in *Proceedings of the International Laser Safety Conference*, vol. 2005, no. 1, March 2005.
 - [27] W. J. Marshall, “Update of national standard for the safe use of lasers outdoors,” *Journal of Laser Applications*, vol. 11, no. 5, pp. 234–236, 1999.
 - [28] A. Krizhevsky, I. Sutskever, and E. Geoffrey, “Imagenet classification with deep convolutional neural networks,” *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097–1105, 2012.
 - [29] J. Deng, “Imagenet: a large-scale hierarchical image database,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, IEEE, Miami, FL, USA, June 2009.
 - [30] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
 - [31] Y. Pu, Z. Gan, R. Henao et al., “Variational autoencoder for deep learning of images, labels and captions,” pp. 2352–2360, 2016, <https://arxiv.org/abs/1609.08976>.
 - [32] A. Mohamed, G. E. Dahl, and G. Hinton, “Acoustic modeling using deep belief networks,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 1, pp. 14–22, 2011.
 - [33] W. Rawat and Z. Wang, “Deep convolutional neural networks for image classification: a comprehensive review,” *Neural Computation*, vol. 29, no. 9, pp. 2352–2449, 2017.
 - [34] Y. LeCun, “Handwritten digit recognition with a back-propagation network,” *Advances in Neural Information Processing Systems*, vol. 2, pp. 396–404, 1990.
 - [35] W. Liu, “Large-margin soft max loss for convolutional neural networks,” *ICML*, vol. 2, no. 3, p. 7, 2016.
 - [36] C. Szegedy, “Going deeper with convolutions,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–9, Boston, MA, USA, June 2015.
 - [37] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” 2014, <https://arxiv.org/abs/1409.1556>.
 - [38] J. Shijie, “Research on data augmentation for image classification based on convolution neural networks,” in *Proceedings*

- of the 2017 Chinese automation congress (CAC), pp. 4165–4170, IEEE, Jinan, China, October 2017.
- [39] X. Glorot, B. Antoine, and Y. Bengio, “Deep sparse rectifier neural networks,” in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pp. 315–323, Lauderdale, FL, USA, April 2011.
 - [40] H. F. Yu, F. L. Huang, and C. J. Lin, “Dual coordinate descent methods for logistic regression and maximum entropy models,” *Machine Learning*, vol. 85, no. 1-2, pp. 41–75, 2011.
 - [41] S. J. Pan and Q. Yang, “A survey on transfer learning,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2009.
 - [42] Z. Xuegong, “Introduction to statistical learning theory and support vector machines,” *Acta Automatica Sinica*, vol. 26, no. 1, pp. 32–42, 2000.
 - [43] R. Hecht-Nielsen, *Theory of the Backpropagation Neural Network, Neural Networks for Perception*, Academic Press, Cambridge, MA, USA, 1992.
 - [44] X. Chen, L. Qi, Y. Yang, Q. Luo et al., “Video-based detection infrastructure enhancement for automated ship recognition and behavior analysis,” *Journal of Advanced Transportation*, vol. 2020, Article ID 7194342, 12 pages, 2020.
 - [45] X. Chen, Z. Li, Y. Yang, L. Qi, and R. Ke, “High-resolution vehicle trajectory extraction and denoising from aerial videos,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 5, pp. 3190–3202, 2021.