Advances in Engineering Research, volume 184
2nd International Conference on Sustainable Energy, Environment and Information Engineering (SEEIE 2019)

Research on Vehicle Classification and Recognition Method Based on Vehicle Acoustic Signal CNN Analysis

Zhangli Lan[1], Yuxin Zhang[1,*], Juan Cao[1], Ranran Tang[2], Liyun Tan[2] and Fang Liu[1]
[1]School of Information Science and Engineering, Chongqing Jiaotong University, Chongqing 400074, China
[2]Chongqing Municipal Highway Bureau, Chongqing 401147, China
*Corresponding author

***Abstract*—The present "shallow classification model" have shortcomings on modeling and representation ability, feature extraction, classification performance and so on. This study aims to improve the typical LeNet-5 convolution neural network and obtain three kinds of CNN structures to realize the classification of large and small vehicles. Firstly, we extracted the MFCC feature of vehicle acoustic signals; then took the feature signals as training samples; lastly adjusted the study rate, convolution kernel size and quantity in accordance with experiment and obtained the results. The experimental results indicate that the improved CNN model is better than the traditional machine learning method; and the classification performance of the improved CNN model is improved with the increase of data volume, and the accuracy of the test samples is 96.8%.**

Keywords—intelligent transportation; vehicle classification recognition; vehicle acoustic signal; feature extraction; deep learning; convolution neural network

I. INTRODUCTION

As an important part of the field of intelligent transportation, automatic classification and recognition of vehicle type plays an important role in the system of automatic collection toll on the highway, car park management, highway traffic flow statistics and analysis and so on.

Classification and recognition methods of vehicle type for vehicle sound and vibration signals have been widely studied at home and abroad. Shuangwei Wang and others collected the ground vibration signal and the external noise signal generated by two kinds of single vehicle models for the AR parameter model, and used the AR model parameter of the vehicle external noise as the characteristic parameter of vehicle classification and recognition. The classification results were up to 80% and more [1]. Guilin Liu, Xianglong Luo and Xialin Ma selected wavelet packet transform, empirical mode decomposition, spectrum analysis and other methods to process vehicle acoustic signals, and adopted SVM to classify and recognize targets[2-4]. William and others adopted wireless sensors to capture the signals of ground vehicles for military, extract harmonic amplitude estimation of harmonic characteristics of acoustic signals in time domain, and use multi-layer feedforward neural network to classify and recognize vehicles[5]. Jinghua Li and others have studied acoustic signal of a motor vehicle. They adopted the wavelet decomposition to obtain the energy of acoustic signals on different scales, which served as eigenvector. And A K-nearest neighbor classifier model and an improved BP neural network are established for target classification [6]. The above methods are based on "shallow model" to realize vehicle classification and recognition by way of vehicle acoustic signal. But the modeling and representation ability is limited, and there is still room for improvement in feature extraction, classification performance and so on.

CNN is an artificial neural network model that is multi-stage, global and trainable, which can process the abstract, essential and high-order characteristics from a small amount of pre-processing and even the original data[7]. Qing Hu and others used the convolution and de-sampling operation of CNN to pre-process the speaker's speech signal. Then they extracted the MFCC feature parameters, and used the classical Universal Background Model to modeling the speaker's cognition model [8]. Abdel-Hamid and others combined with the model Hidden Mal. Cove to establish a model which recognizes speech by CNN and carried out experiments on the standard TIMIT speech database. The experimental results show that the error rate of the model is 10% lower than that of the conventional neural network model with the same number of hidden layers and weights[9].

The complexity of the vehicle acoustic signal makes its signal expression problem possible to make use of highly abstract provided by deep learning. And the convolution and pooling operation of CNN can effectively express and process some typical features of audio signal hidden in frequency domain. This paper proposes a method of vehicle classification and recognition based on CNN analysis of vehicle acoustic signal, and also designs three kinds of CNN models to realize the classification of large and small cars.

II. FEATURE EXTRACTION OF VEHICLE ACOUSTIC SIGNAL

The Mel frequency cepstrum coefficient (MFCC) is similar to the human ear, which can distinguish a big car from a small one according to the sound. Therefore, MFCC is used as the characteristic expression of the vehicle acoustic signal. It effectively combines the auditory perception characteristics of human ears and the mechanism of speech generation, and carries out nonlinear mapping of the original frequency (Hz) of

the sound signal. The extraction process of the MFCC is shown in Figure 1.

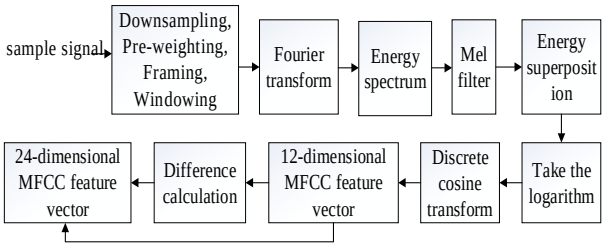

FIGURE I. VEHICLE ACOUSTIC SIGNAL MFCC EXTRACTION PROCESS

After the vehicle sound signal is processed by the above steps, the 24-dimensional MFCC features of each frame signal are obtained. Each MFCC in series produces a sample signal Characteristics. Figure 2 is the waveform in time domain and MFCC characteristic diagram of a small car sample signal. Figure 3 is waveform in the time domain and MFCC characteristic diagram of a large vehicle sample signal.

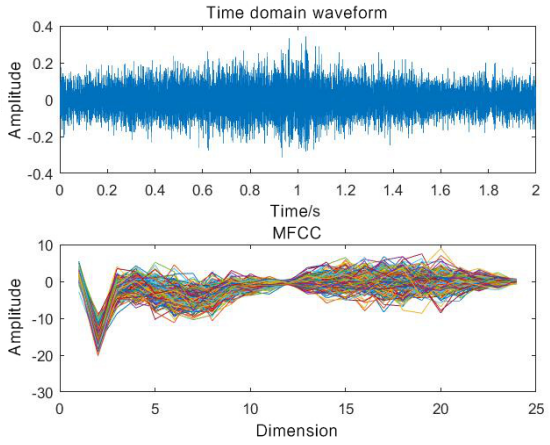

FIGURE II. TIME DOMAIN WAVEFORM AND MFCC CHARACTERISTIC OF A SMALL CAR SAMPLE SIGNAL

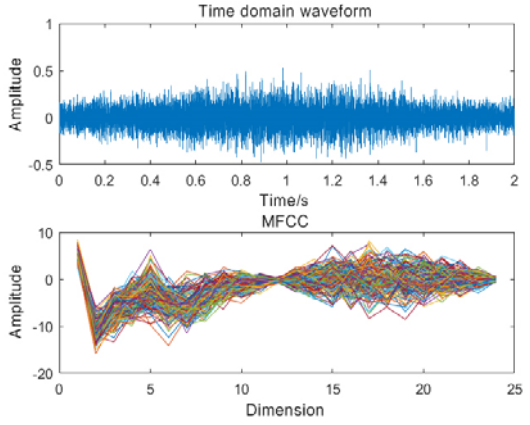

FIGURE III. TIME DOMAIN WAVEFORM AND MFCC CHARACTERISTIC OF A LARGE CAR SAMPLE SIGNAL

III. LENET-5 NETWORK AND ITS IMPROVEMENT

A. LeNet-5 Network

In machine learning, CNN is an in-depth learning neural network model derived from artificial neural networks. CNN-specific convolution and pooling operations can effectively express and process some of the typical features hidden in the frequency domain of audio signals [10-11]. But use the sample model which created in this research, data set 1 and data set 2 train LeNet-5 network, and find that the model does not converge, the loss function value is very high, and has no classification ability. After analysis, the following reasons are summarized:

a) The characteristic data of the vehicle sound signal is about 5 times that of the hand-written digital sample, and the feature type obtained by extracting the original convolution kernel number is too small to be insufficient to effectively identify the vehicle type.

b) The feature number of sample data is different from the pixel value of handwritten digital picture. The super-parameters such as the dimension of input data and the depth of network structure need to be determined.

c) LeNet-5 has no anti-fitting technique.

d) There are 10 types of datasets used by LeNet-5, and there are only 2 types of models in this article.

B. LeNet-5 Network Improvement Process

*a) Through theoretical and experimental comparison, the activation function is ReLU, loss function is cross entropy function, the optimizer is the number of adadelta, convolution kernels is 32, the convolution size is 5*5. Because of the complexity of the model, the depth neural network is prone to over-fitting in the training process. The regularization method Dropout. Dropout is used for setting 0.25 in the pool layer, 0.5 in the fully connected layer, and 1. 1 in batch size.*

b) By setting different learning rates from 0.001 to 1, repeated experiments were carried out, and a comprehensive selection of 0.01 learning rates was used for follow-up experiments.

*c) The larger the convolution kernel size is, the more the convolution layer parameters are, the less the parameters of the full connection hidden layer are, and the number of network parameters is mainly in the full connection layer. The size of convolution nucleus are 3*3, 5*5, 7*3, 7*5, 7*7, respectively. The principle of size selection of convolution kernel is verified, and the 5*5 convolution kernel is selected after comprehensive analysis.*

d) The structure of the network layer consists of two convolution layers and two pooling layers. The original two full connection layers are changed into a fully connected hidden layer, with 128 neurons, followed by one output layer and two neurons. The output layer uses softmax activation function. So far, we have completed the preliminary improvement of LeNet-5, and named the improved LeNet-5 network as the total number of CNN1, layers of 6 layers.

e) Considering that the number of neurons in the full connection layer is usually difficult to determine according to experience, but the number of neurons in the hidden layer has a great impact on the performance of the model. Therefore, the full connection hidden layer of CNN1 is improved, and the full connection hidden layer is changed into a global average pooling layer. Finally, the output layer remains unchanged. So the improved convolutional neural network CNN2, has two convolutional layers in all, and it is proved that there are two convolutional layers in the improved convolutional neural network CNN2, it has 2 maximum pool layers, 1 full Local average pool layer, 1 output layer, the total number of layers is 6.

*f) Considering the change of a global average pool layer into a fully connected layer, the number of features decreases too quickly, which may lead to insufficient feature depth and affect the classification results. Therefore, a convolution layer is added before the global mean pool layer, and the convolution kernel size is 3*3, and the number of convolution kernels in this layer is determined to be 16. So far, the improved convolution neural network CNN3, in this paper has three convolution layers in total. It has 2 maximum pool layer, 1 global average pool layer, 1 output layer, 7 layers.*

IV. EXPERIMENT

A. *Experiment Data Feature Set*

The experiment data is collected acoustical signals of rolling road rumble stripes during vehicles' running by the recording software under the condition that the sample vehicle speed is in the normal range and the speed distribution is the same at where the road rumble stripes are set on the real pavement. The data is mono in WAV format, and the sample rate is 48KHz. Based on the analysis of the samples, matlab2015b is used to preprocess and feature the signals of small and large cars, and to label each sample corresponding category to establish the feature set required for the experiment.

There are 2 experiment data feature sets containing 2806 data samples. The sample size and labels for each data set are shown in Table I.

TABLE I. COMPOSITION OF DATASET

	Total sample	**Data set 1**	**Data set2**	**Label**
Large car	1448	486	962	0
Small car	1358	465	893	1
Total sample	2806	951	1855	

B. *Experimental results and analysis*

The data set 2 is trained and used for classification effect test. The learning rate is set to 0.01, the batch size to 16, and the iterations to 100 and 200 respectively to verify the performance of the three networks. The results are as shown in Figure 4 – Figure 7 and Table II.

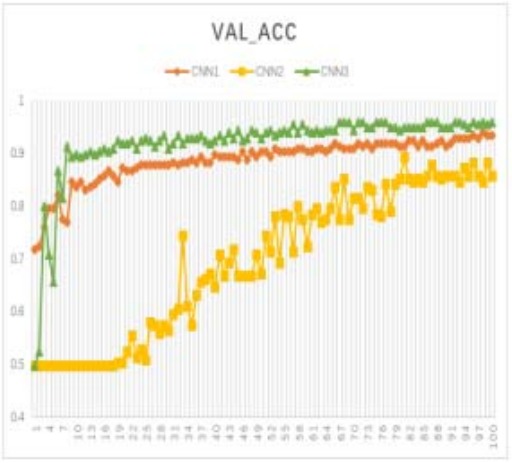

FIGURE IV. THE COMPARISON OF EXPERIMENTAL ACCURACY FOR 100 ITERATIONS OF THREE NETWORKS

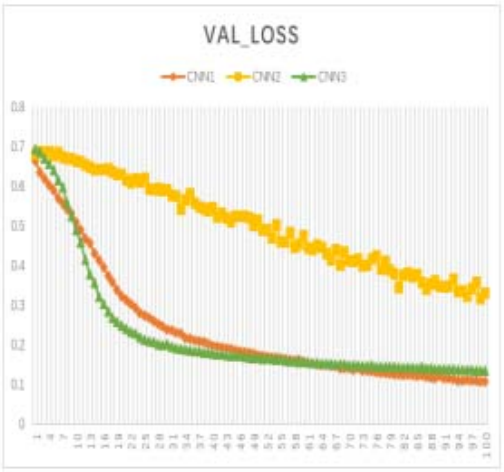

FIGURE V. THE COMPARISON OF EXPERIMENTAL LOSS VALUES FOR 100 ITERATIONS OF THREE NETWORKS

FIGURE VI. THE COMPARISON OF EXPERIMENTAL ACCURACY FOR 200 ITERATIONS OF THREE NETWORKS

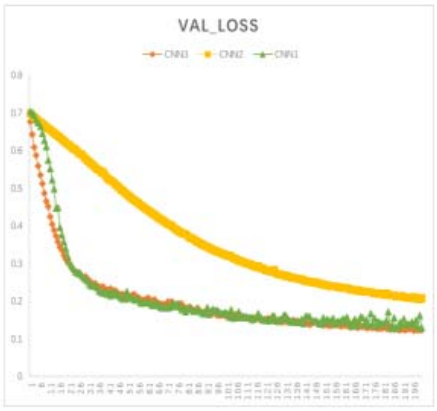

FIGURE VII. THE COMPARISON OF EXPERIMENTAL LOSS VALUES FOR 200 ITERATIONS OF THREE NETWORKS

TABLE II. RESULTS OF THE THREE IMPROVED CONVOLUTIONAL NEURAL NETWORKS ON DATASET 2

	CNN1		CNN2		CNN3	
epochs	100	200	100	200	100	200
val_acc（%）	94.2	95.9	85.6	92.3	94.8	95.4
val_loss	0.11	0.12	0.33	0.24	0.15	0.14
Time（s）	156	307	136	272	137	269

TABLE III. THE COMPARISON OF ALGORITHM ACCURACY(%)

	KNN	RBF	Linear	PCA+RBF	CNN1	CNN2	CNN3
Data set 1	77.5	51.2	88.4	88.7	94.1	92.6	95.6
Data set 2	69.3	53.4	81.2	81.3	96.5	93.8	96.8

In Figure 4 - Figure 7, val_acc is the accuracy of the test set, and val_loss is the loss function value of the test set. It can be seen from Figure 4 and Figure 5 that when the network iterates 100 times, CNN2 is in a state of severe under-fittingwhen the number of iterations is increased, CNN2 has the worst learning ability, and CNN3 is slightly better than CNN1. Figure 6 and Figure 7 show that when the iterations are 200 times, CNN1, CNN2 and CNN3 are in a steady learning state. CNN2 has the worst learning ability, and CNN3 is slightly better than CNN1.

Table II shows that in the same data set, with the increase of iterations, the performance of the designed three CNN networks is getting better and better. The accuracy of the networks is improved and the value of the loss function decreasing. Among them, CNN3 has the best performance.

In order to further verify the performance of the three CNN networks, the iterations is set to 300. The three CNN networks are compared on different data volumes with the traditional vehicle classification results with the value of val_acc as the evaluation index. The results are shown in Table III.

It can be seen from Table III that the traditional machine learning algorithm KNN has a general classification effect on the data set of this paper, and SVM a good classification effect, but both are not as good as the improved three CNN models. With the increase of data volume, the classification performance of convolutional neural network is improved, while that of the traditional shallow model decreases, which indicates that the convolutional neural network has certain advantages in processing large data sets.

V. CONCLUSION

This paper proposes a vehicle recognition method of CNN analysis based on vehicle acoustic signal against the shortages of traditional method of vehicle classification and recognition. Three kinds of CNN structures are designed by optimizing parameters and improving CNN structure. The experiment shows that the classification accuracy of the improved CNN model is significantly better than that of the traditional machine learning method, and CNN classification performance has improved with the increase of data volume.

ACKNOWLEDGEMENT

This research was supported by Science and technology research project of Chongqing education commission (KJQN20180716).

REFERENCES

[1] Shuangwei Wang, Qiang Chen, Jiang Li, Hongfeng Wei, Liping Du and Lihua Zhao.A study on the characteristics of Sound and Vibration signals of different vehicle models[J]. Acoustics Technology,2007,26(03):460-463.

[2] Guilin Liu, Xiangwei Kong, Hang Liu. Research on vehicle Detection and recognition algorithm based on Wireless Sensor Network[J].Sensors and microsystems,2010,29(02)：9-12.

[3] Xianglong Luo, Guohong Niu, Qianjiao Wu,and Ruoyu Pan.Recognition of vehicle Audio Frequency based on empirical Mode decomposition and support Vector Machine [J].Applied acoustics,2010,29(3):178-183.

[4] Xialin Ma, Ming Cai, Jianli Ding. Research on vehicle Audio recognition based on Spectrum Analysis and support Vector Machine[J].Applied Acoustics,2014(04):13-17.

[5] William P E, Hoffman M W.Classification of Military Ground Vehicles Using Time Domain Harmonics'Amplitudes[J].IEEE Transactions on Instrumentation & Measurement,2011,60(11):3720-3731.

[6] Jinghua Li, Yifeng Zhao, Jiadong Xu.Vehicle recognition based on wavelet energy and neural network classifier [J].Journal of China Highway,2007,20(03):97-102.

[7] Qingqing Zhang, Yong Liu, Jielin Pan ang Yonghong Yan. Continuous speech recognition based on convolution neural network [J]. Journal of engineering science,2015,37(9):1212-1217.

[8] Qing Hu, Benyong Liu. Speaker recognition algorithm based on Convolutional Neural Network [J]. Computer application,2016,36(s1):79-81.

[9] Abdel-Hamid O, Mohamed A R, Jiang H, Deng L,Penn G and Yu D.Convolutional Neural Networks for Speech Recognition[J].IEEE/ ACM Transactions on Audio Speech & Language Processing, 2014,22(10):1533-1545.

[10] Abdel-Hamid O, Li D, Dong Y. Exploring Convolutional Neural Network Structures and Optimization Techniques for Speech Recognition[C]//interspeech,2013（1）：1173-1175.

[11] Li Deng, Jinyu Li, Juiting Huang, Kaisheng Yao, Dong Yu,Seide Frank,et al.Recent advances in deep learning for speech research at Microsoft[C]//In Proceedings of International Conference on Acoustics Speech and Singal Processing(ICASSP),2013(16):8604-8608.