Research on Pavement Type Recognition Algorithm Based on Improved ResNeXt-50

Farong Kou, Liujie Xu, Jie Ren, Haiqi Wang, Ling Hu, Yuan Li

***Abstract*—This paper obtains accurately the pavement types and serves the intelligent vehicle suspension control system. An improved pavement types identification algorithm with improved ResNeXt-50 is proposed in this paper. Firstly, improved ResNeXt-50 is established as the backbone network. Secondly, the lightweight CBAM attention mechanism and improved ASPP module are used. They extract the contextual information and multi-scale detail features of the input image to solve the problem that the network loses the detail information of the input image. Finally, the H-Swish activation function is used in the deep network to reduce the number of parameters and training time. The experimental analysis shows that the identification accuracy of the improved network model for various pavement types respectively reaches 93.8%, 92.8%, 95.9%, 94.7%, 91.2%, and 99.2%. It is a noteworthy improvement over the traditional BP neural network.**

***Index Terms*—ResNeXt-50, pavement types, improvement of ASPP, CBAM attentional mechanism**

I. INTRODUCTION

With the development and application of artificial intelligence, intelligent control of vehicles has become a research hotspot. Pavement types identification as the first condition for controlling the vehicles suspension system. Accurate recognition results are of great significance for intelligent vehicle control. With the development of environmental perception and attitude cognition of intelligent vehicles [1], there are two methods about the pavement types identification, direct method and indirect method. The direct method mainly relies on laser radar, LIDAR and other on-board sensors [2][3] to directly obtain data and analyze the pavement surface state. And the indirect method is mainly through the establishment of accurate active suspension dynamics model to identify the current pavement types.

Wang [4] proposed a pavement classification model based on structural re-parameterization and adaptive attention. The model can quickly and accurately screen complex pavements such as asphalt, concrete, snowy pavement, frozen pavement and sand pavement. Neupane [5] proposed a heuristic LIDAR-based pavement types detection method. He identified the pavement types by the mean and variance of the laser reflected intensity. The method is mainly used for asphalt pavement. Liu [6] proposed adaptive attenuation traceless Kalman filtering algorithm for observing the pavement adhesion coefficient. The algorithm can be adapted to the estimation of many types of pavement. Xiaoyan Lu [7] proposed a multi-task road extraction framework that simultaneously extracted pavement surface, center line and edge and improved pavement information integrity. Siyun Chen [8] realized the detection of pavement elements in any direction from a cloud of moving laser scanning points and achieved high precision detection. Sahu [9] built the RCNet pavement identification model, which can identify different states of pavement such as frozen pavement and water-covered pavement. Chen [10] fused model and data driven approach for pavement type identification.

Since the LiDAR used in the direct method is relatively expensive. The indirect method can only identify the pavement types under the current tires. It has an obvious lag in the identification of the pavement types. Based on these, a pavement types identification network is proposed based on an improved ResNeXt-50 network. The network comprises a lightweight CBAM attention mechanism and improved ASPP module to realize the identification of the pavement types by using multi-scale detail features.

Manuscript received July 12, 2024; revised January 14, 2025.

This work was supported in part by the China National Natural Science Foundation under Grant 51775426.

Farong Kou is a master's graduate from Northwestern Polytechnical University, Xi'an, 710054 China (phone: 18729095156; e-mail: 342546738@qq.com).

Liujie Xu is a postgraduate student of Xi'an University of Science and Technology, Xi'an, 710054 China. (e-mail: 2030047588@qq.com).

Jie Ren is a postgraduate student of Xi'an University of Science and Technology, Xi'an, 710054 China. (e-mail: 978891098@qq.com).

Haiqi Wang is a postgraduate student of Xi'an University of Science and Technology, Xi'an, 710054 China. (e-mail: 923995170@qq.com).

Ling Hu is a postgraduate student of Xi'an University of Science and Technology, Xi'an, 710054 China. (e-mail: 1603586383@qq.com).

Yuan Li is a postgraduate student of Xi'an University of Science and Technology, Xi'an, 710054 China. (e-mail: 3124668272@qq.com).

II. PAVEMENT IMAGE ACQUISITION AND PRODUCTION OF DATASETS

Based on six common pavement types [11], the opened source dataset RSCD of Tsinghua University team is selected. The vehicle for collecting as shown in Fig. 1. The selected pavement types are high quality asphalt, medium asphalt, poor quality asphalt, concrete, gravel and dirt pavement as shown in Fig. 2. Images of each pavement type contains 10,000 samples. The samples are divided into training set, validation set and test set in the ratio of 8:1:1, as shown in Table I.

TABLE I
DATA SET COMPOSITION

type of pavement	filename	(train/val/test)
high quality asphalt	asphalt_smooth	8000/1000/1000
medium asphalt	asphalt_slight	8000/1000/1000
poor quality asphalt	asphalt_severe	8000/1000/1000
concrete	concrete	8000/1000/1000
gravel	gravel	8000/1000/1000
dirt	mud	8000/1000/1000

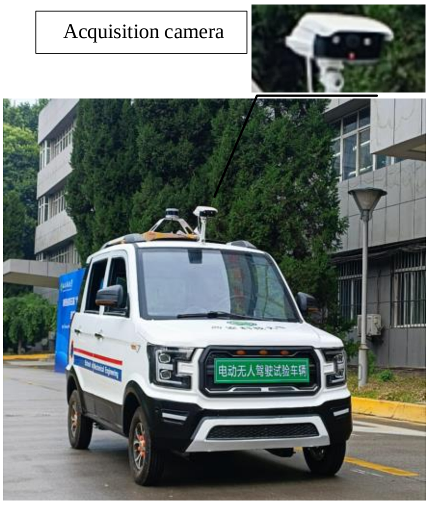

Fig. 1. Pavement image acquisition vehicle

Fig. 2. Pavement image. The six pavement types are used in the dataset.

III. CBAM ATTENTION MECHANISM AND IMPROVED ASPP MODULE

To settle the trouble of gradient vanishing and gradient explosion that appeares during the training process about deep neural networks. He [12] proposed Residual Network (ResNet). The residual block used jump connections based on ordinary CNN to achieve the purpose of enabling the network to learn the residuals. And the residual block composition is shown in Fig. 3. The dimensionality reduction operation is performed by 1×1 convolution kernel to reduce the number of channels of the input feature matrix from 256 to 128. Then 32 groups convolutions are aggregated and spliced after feature extraction using four 3×3 convolution kernels. Finally, the input is upscaled and summed with the outputs by a 1×1 convolution kernel. The residual block is calculated as shown in Equation 1.

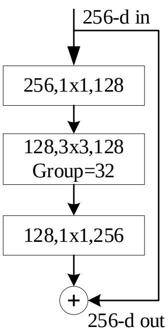

Fig. 3. Residual structure diagram

$$y = x + \sum_{i=1}^{C} T_i(x) \tag{1}$$

where: x is the input, y is the output, C is the base, express the size of the set of aggregation transformations, and $T_i(x)$ is an arbitrary mapping function for the pair to x.

In deep learning, the attention mechanism is added for effective extraction of feature maps. By mimicking the way of the human brain thinks about processing visual information. The mechanism is possible to process visual information more efficiently. The mechanism could focus attention quickly on important regions and ignore other irrelevant background regions [13]. Most of the current attention meshanism are divided by space attention and passageway attention. The paper is introduced a lightweight multi-channel convolutional attention module CBAM (Convolutional Block Attention Module). CBAM contains space attention and passageway attention mechanisms. The composition of the CBAM network is shown in Fig. 4(a).

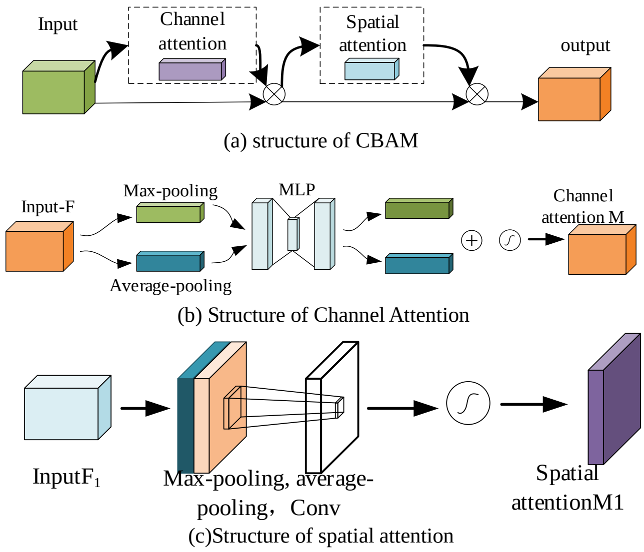

Fig. 4. CBAM attention structure diagram

The spatial information of the pavement feature mape is firstly aggregated by mean-pool and max-pool operations. It generates two different spatial feature maps: mean-pool feature map and max-pool feature map. Then the two feature maps are shared to the multilayer perceptron network (MLP). And the features output from the MLP are subjected to the add and sigmoid activation operations to generate channel attention feature maps M [14] in Fig. 4(b). The channel attention feature map M calculation formula is shown in Equation 2.

$$M(F) = \sigma(MLP(avgpool(F)) + MLP(\max pool(F))) \tag{2}$$

Where: $M(F)$ denotes the channel attention feature map, σ denotes the sigmoid activation function.

A multiplication operation is done between the feature map M and the initial input feature map. Then the input features of the pavement image required by the spatial attention module. The input feature maps in the spatial attention channel are subjected to max-pooling and average-pooling, followed by channel splicing and convolution operations. Degradation and sigmoid activation to generate the feature map M1 is shown in Fig. 4(c). The spatial attention feature map formula is shown in Equation 3.

$$M1(F_1) = \sigma(f^{7\times 7}([avgpool(F_1); \max pool(F_1)])) \quad (3)$$

Where: $M1(F_1)$ denotes the spatial attention feature map, $f^{7\times 7}$ denotes the 7×7 convolutional layer, $F_1 = M(F) \otimes F$.

The ASPP [15] multi-scale feature extraction module increases the sensing field and improves the network power about perceiving targets at different scales without increasing the computational effort. The module adds cavity convolutions based on Spatial Pyramid Pooling (SPP). The module consists of a series of expansion convolutions with different expansion rates rates and global-pooling layers. This module can complete convolution and mean-pool operations on the input image. After fusion, the multiscale description of image context information is realized. But the pooling operation will reduce the resolution about the import picture while increasing the sensory field. The input image will lose the key information. The input image after pooling will not be able to be restored after up-sampling, which is a disadvantage that ultimately limits the accuracy of the classification. The improved ASPP structure consists of three parts: 1×1 convolution, pooling pyramid and adaptive mean pooling. The 1×1 convolution is performed for dimensionality reduction. The adaptive mean-pooling is performed for global feature extraction for each channel. In the pooling pyramid part, the expansion factor is set separately, and the corresponding null convolution layer is superimposed to extract features at different scales. After that, the feature map is 3×3 convolved and then spliced with the original input image to fully obtain the detailed information missed due to the null convolution. Finally, the images of the five channels are concatenated to get the output in Fig. 5.

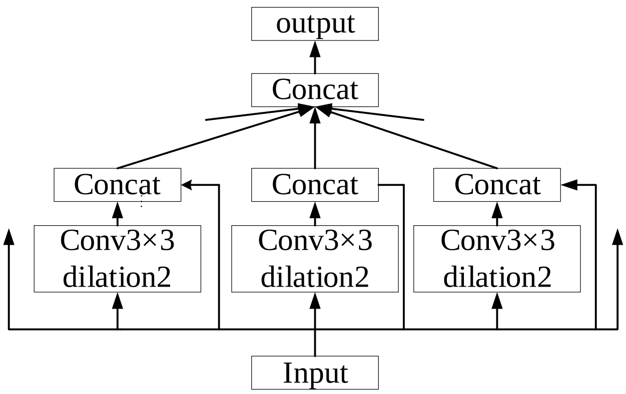

Fig. 5. Structure of the improved ASPP. The leftmost channel is for 1×1 convolution, the middle three channels are for the improved part, and the rightmost channel is for up-sampling.

$$H-Swish(x)=x\cdot(\mathrm{ReLU6}(x+3))/6=\begin{cases}0, x\le -3\\ x, x\ge 3\\ x\cdot(x+3)/6, \text{others}\end{cases} \quad (4)$$

The H-Swish function [16] is an approximate activation function proposed in MobileNetV3. Compared to the ReLU activation function and derived activation functions, the H-Swish activation function can increase the accuracy of the neural network well in deep networks. H-Swish formula is shown in Equation 4.

To calculate the dissimilarity between the predicted value and the real value of the improved network. The loss function defined as the cross-entromy loss function. For the multiclassification problem, the softmax activation function is used to normalize the cross-entromy loss ability formula and softmax activation function [17]. The formula is shown in Equation 5 and 6.

$$L = \frac{1}{N}\sum_i L_i = -\frac{1}{N}\sum_{i=0}^{N-1}\sum_{c=1}^{M-1} y_{ic}\log(p_{ic}) \quad (5)$$

Where: M is the total number of pavement class categories, N is the size of the constructed dataset, p_{ic} is the probability of predicting sample i to be in category c, and y_{ic} is the actual probability with a value of 0 or 1.

$$Softmax(p_i) = \frac{e^{p_i}}{\sum_{i=1}^{M} e^{p_i}} \quad (6)$$

Where: M is the total number of pavement class categories, and p_i is the class output probability of the pavement.

IV. Improved ResNeXt-50 Network

RestNeXt-50 [18] is a stackable deep network model, unlike traditional CNN networks. ResNeXt-50 is introduced the split-transform-aggregate strategy of the Inception family of networks. It transforms a single convolution into multiple convolutions with the same topology in multiple branches and reduces hyperparameters. The improved ResNeXt-50 network consists of 5 Conv layers, 2 CBAM layers, 1 improved ASPP layer, 1 softmax function layer and 2 pooling layers. The structure of improved ResNeXt-50 in Table II, and the composition of network model is shown in Fig. 6.

TABLE II
IMPROVED RESNEXT-50 NETWORK MODEL STRUCTURE TABLE

Layer	output	ResNeXt-50
Conv1	112×112	7×7,64, Stride=2
CBAM1	112×112	3×3, max-pooling, Stride=2
ASPP	56×56	\
Conv2	56×56	$\begin{bmatrix}1\times1 & 128 & \\ 3\times3 & 128 & C=32\\ 1\times1 & 256 & \end{bmatrix}\times 3$
Conv3	28×28	$\begin{bmatrix}1\times1 & 256 & \\ 3\times3 & 256 & C=32\\ 1\times1 & 512 & \end{bmatrix}\times 4$
Conv4	14×14	$\begin{bmatrix}1\times1 & 512 & \\ 3\times3 & 512 & C=32\\ 1\times1 & 1024 & \end{bmatrix}\times 6$
Conv5	7×7	$\begin{bmatrix}1\times1 & 1024 & \\ 3\times3 & 1024 & C=32\\ 1\times1 & 2048 & \end{bmatrix}\times 3$
CBAM2	7×7	average-pooling, Softmax

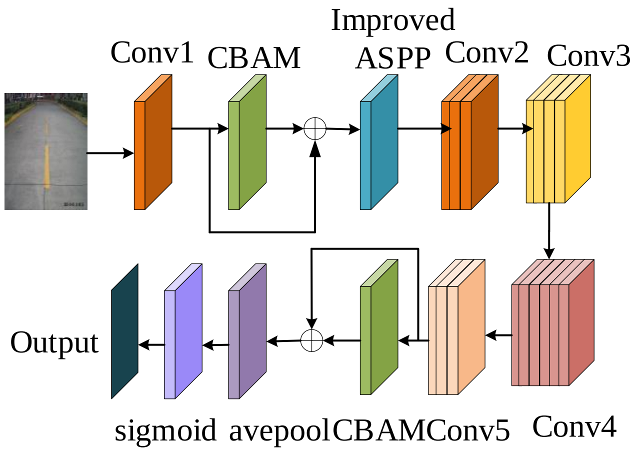

Fig. 6. Structure of improved ResNeXt-50 network. This figure corresponds to Table II and represents the specific identification process for the improved ResNeXt-50 network.

V. Improved ResNeXt-50 Network Validation

In an effort to make full use of the dataset intel, that input images are firstly subjected to pre-processing operations: ① Convert all pavement images into three-channel RGB images of size 224×224. ② Carry out data enhancement operations such as cropping and panning on all pavement images. ③ The pavement images dataset is standardized and disrupted. The improved ResNeXt-50 is trained and verified for reliability according to the Owned dataset. The configuration of the test environment used for the pavement recognition algorithm in this paper. They are shown in Table III.

TABLE III
ENVIRONMENT CONFIGURATION TABLE

Operating system	CPU	GPU	environment
Windows 11	R7-5800H	NVIDIA GTX3060	PyTorch

The improved network reaches a stable convergence state after 150 rounds of training, no overfitting occurs. The precision and loss curves are shown in Fig. 7. The optimal model is saved for testing on the test collections, and the confusion matrix of the experiment outcome is shown in Fig. 8.

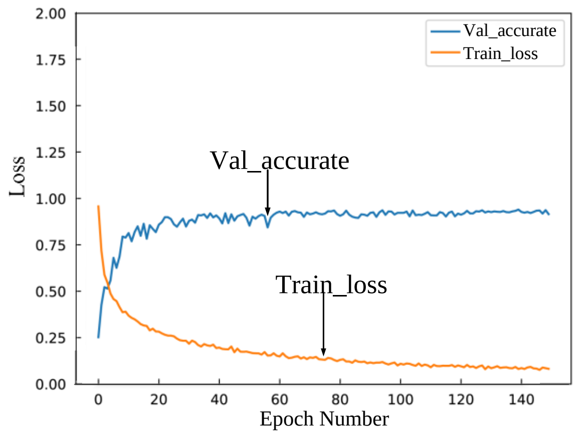

Fig. 7. Accurate and loss curves

Recall, precision and F1 score are used as the judgment indexes for pavement types identification. That calculation formula is shown in Equation 7 and the evaluation outcomes are shown in Fig. 9.

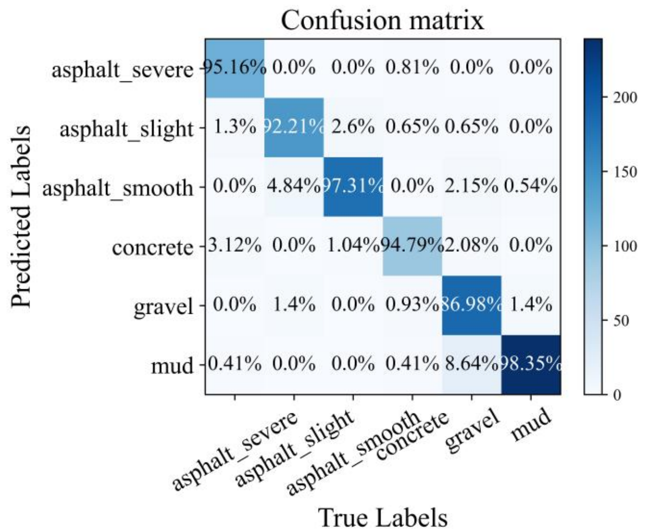

Fig. 8. Confusion matrix of Improved Resnext-50

$$
\begin{aligned}
pre &= \frac{TP_i}{TP_i + FP_i} \\
rec &= \frac{TP_i}{TP_i + FN_i} \\
F1 &= \frac{2\,pre \times rec}{pre + rec}
\end{aligned}
\tag{7}
$$

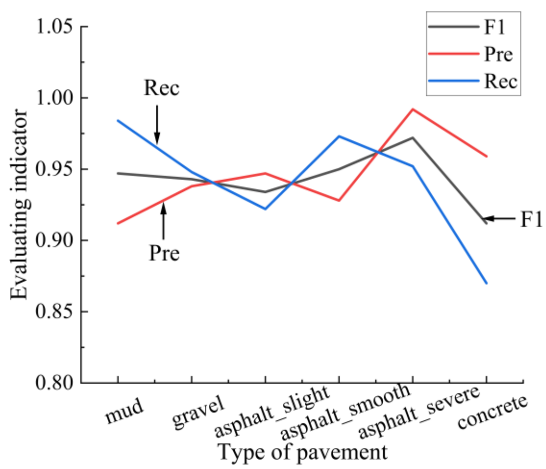

Fig. 9. Pre, rec and F1 score. F1 denotes the reconciled mean of pre and rec, the precision-curve describes the proportion of correctly predicted outcomes, and rec determined the proportion of positive samples properly discriminated by the model.

Comparison of the effect of adding different modules of the ResNeXt-50 network on the classification effect through ablation experiments. Denote the improved model used as the CM-As-ResNeXt-50 model. The model with the addition of the CBAM attention mechanism as CB-ResNeXt-50. The model with the addition of the improved ASPP module as the As-ResNeXt-50 model. The Loss curves and Accuracy curves of the four methods are shown in Fig. 10 and Fig. 11.

From the Loss curve graph, the improved ResNeXt-50 network has a larger decreasing trend at the beginning of training. But the convergence speed is slower by the accuracy graph. The improved network has improved the precision of classification of all pavements. The precision of recognition of medium asphalt pavement have been improved more.

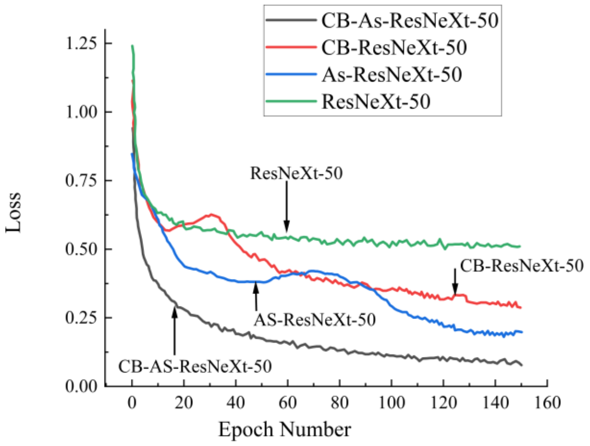

Fig. 10. Loss curve of ablation experiment. The grey, red, blue and green curves represent the variation curves of loss values for different networks respectively

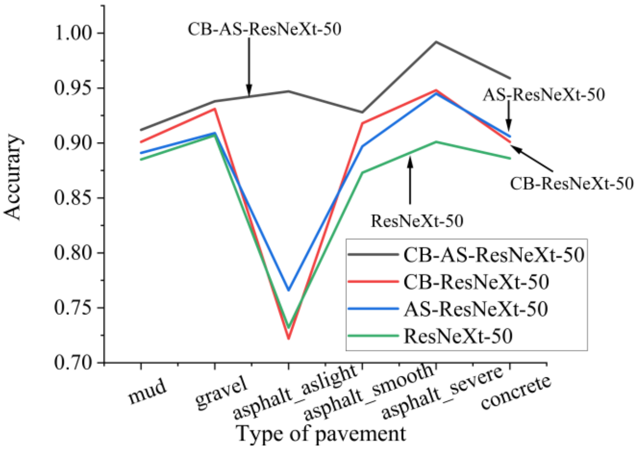

Fig. 11. Loss curve of ablation experiment. The grey, red, blue and green curves represent the precision rate variation curves for different networks, respectively.

In this paper, two comparative experiments are designed to obtain the confusion matrix and accuracy respectively. The results are shown in Fig. 12, Fig. 13 and Fig. 14.

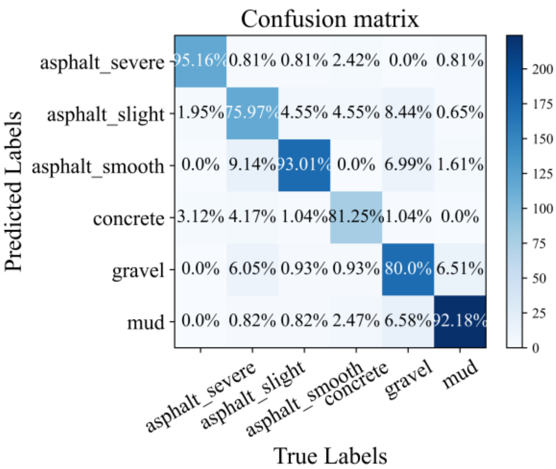

Fig. 12. Confusion matrix of EfficientNet

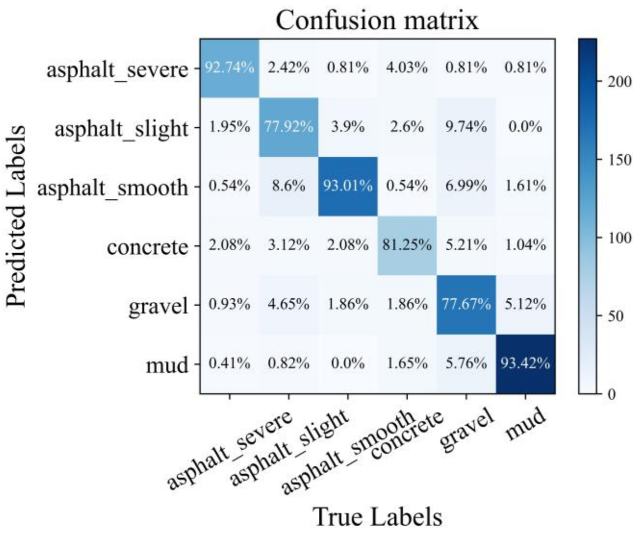

Fig. 13. Confusion matrix of GoogLeNet

The results show that the improved ResNeXt-50 has improved the precision of pavement types identification compared with EfficientNet [19] and GoogLeNet [20] from confusion matrix and accuracy.

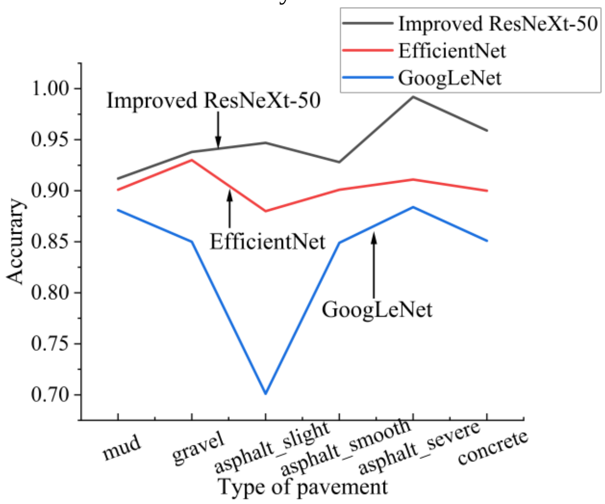

Fig. 14. Comparison experiment. Compare the classification results of the improved ResNeXt-50 network with EfficientNet and GoogLeNet. The black curve is the precision of the improved network. The red curve is the precision of EfficientNet. The blue curve is the precision of GoogLeNet.

VI. Conclusion

The paper constructs a real driving scene dataset, covering 6 common pavement types images. Based on the improved ASPP module and CBAM attention mechanism, they are applied to ResNeXt-50 classification network. They improve the generalization ability and learning ability of network. After trained and tested in PyTorch framework, the improved ResNeXt-50 network model achieved 93.8%, 92.8%, 95.9%, 94.7%, 91.2% and 99.2% precision in identifying high quality asphalt, medium asphalt, poor quality asphalt, concrete, gravel and dirt pavement, respectively. Compared with the traditional ResNeXt network. The classification precision of medium asphalt pavement improved by nearly 19%. Compared with other network models, the classification precision of the improved ResNeXt-50 is also the highest. The experimental results prove that it can meet the requirements of intelligent control.

REFERENCES

[1] Yin Huilin, Wu Shu-li, Wang Yawei, et al. Environment perception and situation cognition Model of Intelligent Vehicle based on Hierarchical Factor Graph [J]. Control and Decision, 2019, 35(10): 2528-2534.

[2] Rateke T, Justen K A, Chiarella V F, et al. Passive Vision egion-Based Road Detection: A Literature Review[J]. Acm Computing Surveys, 2019, 52(2): 31.

[3] Kong Wei, Liu Yun, Li Hui, et al. Review of pedestrian trajectory prediction methods based on Deep learning[J]. Control and Decision, 2021, 36(12): 2841-2850.

[4] Wang Xuewei, Li Si-yuan, Liang Xiao et al. Rapid recognition model of complex pavement based on structural reparameterization and adaptive attention [J/OL]. China Highway Journal: 1-18[2024-03-18]

[5] NEUPANE S R, GHA R AIBEH N G. A Heuristics-based Method for Obtaining Road Surface Type Information from Mobile Lidar for Use in Network-level Infrastructure Management[J]. Measurement, 2019, 131: 664－670．

[6] Liu Zhiqiang, Liu Yiqun. Adaptive attenuation Kalman filter estimation of Road adhesion Coefficient[J]. China Journal of Highway and Transportation, 20, 33(07): 176-185.

[7] X. Lu, Y. Zhong, Z. Zheng, et al. Cascaded Multi-Task Road Extraction Network for Road Surface, Centerline, and Edge Extraction[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-14.

[8] S. Chen, Z. Zhang, H. Ma, et al. A Content-Adaptive Hierarchical Deep Learning Model for Detecting Arbitrary-Oriented Road Surface Elements Using MLS Point Clouds[J]. IEEE Transactions on Gesci-ence and Remote Sensing, 2023, 61: 1-16.

[9] D. K. Dewangan, S. P. Sahu. RCNet: road classification convolutional neural networks for intelligent vehicle system[J]. Intelligent Service Robotics, 2021, 14(2): 199-214.

[10] Chen Kai, Shi Shao-yang, Cheng Shan-shan et al. Research on Road Roughness Recognition Algorithm based on RELIEF-RBF [J]. Journal of Automotive Engineering, 2019, 14(01): 49-59.

[11] Kou Farong, Hu Kailun, Chen Ruochen et al. Predictive Control of Active Suspension Model Based on ResNeSt Network Road condition Recognition[J/OL]. Control and Decision 2022. 1392.

[12] He Kaiming, Zhang Xiangyu, Ren Shaoqing, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, N. V., USA, 2016: 770-778.

[13] Ren Huan, Wang Xuguang. Review on attention mechanism [J]. Journal of Computer Applications, 2021, 41(Supp.1): 1-6.

[14] YANG H, GAO C, GE H, et al. Risk assessment of aviation DC series arc based on reconstructed CBAM-CNN[J]. Journal of Power Electronics, 2023, 23(5): 811-820.

[15] CHEN L C，PAPANDREOU G，KOKKINOS I， et al. DeepLab：semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834-848.

[16] Zhu Xuechao, Zhang Fei, Gao Lu et al. Research on Speech recognition based on residual networks and gated Convolutional networks [J]. Computer Engineering and Applications, 2022, 58(07): 185-191.

[17] CAI Qiang, Li Jing, Hao Jiayun. Remote Supervision Relation Extraction Based on Focusing Loss and Residual Network[J]. Computer Engineering, 2019, 45(12): 166-170.

[18] Xie S, Girshick R, Dollár, Piotr, et al. Aggregated Residual Transformations for Deep Neural Networks[J]. IEEE, 2016.

[19] Tan M, Le Q V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks[J]. 2019.

[20] Szegedy C, Liu W, Jia Y, et al. Going Deeper with Convolutions[J]. IEEE Computer Society, 2014.

Farong Kou, born in Jiuquan in 1973, Gansu Province, graduated from the School of Mechanical and Electrical Engineering, Northwestern Polytechnical University, China, with a PhD degree in Mechanical Engineering in 2008. Main research areas are intelligent networked vehicle technology and vehicle dynamics and control. 2003-2009, Xi'an University of Science and Technology (XUST), School of Mechanical Engineering, Teaching Secretary, Lecturer. 2010-2014, Xi'an University of Science and Technology (XUST), School of Mechanical Engineering, Associate Professor, Head of Department of Vehicle Engineering. 2014-2017, Xi'an University of Science and Technology (XUST), School of Mechanical Engineering, Professor, Vice Dean. Professor, Vice Dean. 2018-2022, Xi'an University of Science and Technology, School of Mechanical Engineering, Professor, Doctoral Supervisor, Vice Dean. 2023-present, Xi'an University of Science and Technology, School of Electrical and Control Engineering, Professor, Doctoral Supervisor, Dean.