J. Vis. Commun. Image R. 64 (2019) 102638

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Road surface condition classification using deep learning ☆

Lushan Cheng [a], Xu Zhang [a,*], Jie Shen [b]

[a] School of Mechanical and Automotive Engineering, Shanghai University of Engineering Science, Shanghai 201620, China
[b] Department of Computer and Information Science, University of Michigan-Dearborn, Dearborn, MI 48128, USA

ARTICLE INFO

Article history:
Received 22 June 2019
Revised 6 September 2019
Accepted 7 September 2019
Available online 9 September 2019

Keywords:
Deep learning
Road condition
Activation function
Image recognition
Intelligent driving

ABSTRACT

Traditional image recognition technology currently cannot achieve the fast real-time high-accuracy performance necessary for road recognition in intelligent driving. Deep learning models have been recently emerging as promising tools to achieve this performance. The recognition performance of such models can be boosted using appropriate selection of the activation functions. This paper proposes a deep learning approach for the classification of road surface conditions, and constructs a new activation function based on the rectified linear unit Rectified Linear Units (ReLu) activation function. The experimental results show a classification accuracy of 94.89% on the road state database. Experiments on public datasets demonstrate that the proposed convolutional neural network model with the improved activation function has better generalization and excellent classification performance.

1. Introduction

In recent years, research in autonomous driving has been widely emerging. Automatic driving technologies have a great potential in reducing traffic congestion and accidents, and in reducing energy consumption and protecting the environment. Indeed, image classification plays an important role in the success of automatic driving technologies, and better classification models can effectively improve safety levels for automatic driving. In particular, image classification can be applied in analyzing road scenes, where a road surface can be classified as dry, icy, wet, snowy or muddy. However, the quality of a road surface image can be affected by weather, illumination changes, motion blurring and other adverse events. These real-world variations negatively impact the performance of the traditional image-based classification methods for road surface conditions, and hence lead to unstable accuracies and poor adaptability of these methods. Therefore, better and more robust methods are needed for road surface condition classification.

In the 1960 s, Hubel and Wiesel found that the unique network structure could effectively reduce the complexity of the feedback neural network when studying the neurons used for local sensitivity and direction selection in the cerebral cortex of cats, and then proposed the convolutional neural network (CNN) [1]. Multilayer neural networks [2] were proposed in the 1980 s as learning models for classification and regression problems. LeCun et al. applied convolutional neural networks [3,4] in character recognition, and hence reduced the workload of manual feature extraction. Krizhevsky et al. proposed the classical convolutional neural network structure [5,6], and made important breakthroughs in image recognition tasks. In recent years, convolutional neural networks (CNNs) have been widely used in many fields, and have shown superior performance in problems such as image target detection [7] and classification [8,9]. For example, Kuehnle et al. [10] proposed a classical neural network structure to classify and recognize a pavement state, but the reported accuracy was low. Liu et al. [11] proposed a pavement state recognition method based on color-space features while Zhang et al. [12] proposed an alternative method based on an improved support vector machine(SVM). However, the results of this method were based on a small sample size and an accuracy of merely 85% was obtained in a mixed road condition recognition task.

Deep learning has demonstrated outstanding success in image classification tasks. One key issue to improve deep learning models is the innovative design of novel activation functions [13]. Activation functions allow a deep learning network model to perform non-linear modeling [14]. Indeed, if a deep network contains only linear convolution and fully-connected layers, then this network could only express a linear mapping, and would not be able to effectively analyze and process non-linear data. After adding a non-linear activation function to the network structure, the network can learn realistic layered non-linear mappings.

The optimization of the rectified linear unit (ReLU) activation function [15–17] can not only significantly accelerate the training

☆ This paper has been recommended for acceptance by Maofu Liu.

* Corresponding author.
E-mail address: zxu1116@126.com (X. Zhang).

https://doi.org/10.1016/j.jvcir.2019.102638
1047-3203/

speed of a deep neural network, but also reduce the training error. In this work, we apply a deep learning method for the classification of images of road surfaces. In addition, we study the influence of the activation function selection on image classification accuracy, and propose an activation function of non-linear terms. We carry out experiments on open-source road datasets (Oxford RobotCar Dataset [18] and KITTI Data Set [19]). The experimental results show that our proposed activation function improves the classification accuracy of road surface conditions. Samples of the five categories of the open-source road surface dataset are shown in Fig. 1.

2. Deep learning models

Deep learning models are machine learning models that seek to imitate the human brain. The first deep learning models were proposed by Hinton et al. in 2006 [20], as an extension of the earlier models of artificial neural networks built from multiple layers of perceptrons. A deep learning model typically contains eight or nine layers, or more. Since each node in these layers has several hyperparameters such as the neuron connection weights and the threshold, the network is capable of automatically extracting many complex features.

Traditional image classification models has the problem of difficulty in selecting suitable features, and poor reliability in practical applications. However, with the development of modern image recognition technologies and techniques, a large amount of image data can be processed by a deep learning model. This leads to a reduced training time and an improved reliability in practical applications. Deep learning models can serve the purpose of automatic feature learning. Through non-linear transformations performed by multiple hidden layers, the “low-level” features can be transformed into abstract “high-level” features. Hence, deep learning models can be used to solve complex classification problems.

In this paper, a convolutional neural network (CNN) is used to classify road surface conditions. The CNN architecture and parameters were set, by experimentation, as shown in Table 1 A schematic diagram of the proposed CNN architecture is shown in Fig. 2.

3. Modified ReLU activation function

Activation functions are used in deep learning models to solve the problem of linear inseparability of raw data, and generate better representations of the data information content. The mathematical form of the ReLU activation function is simple:

Fig. 1. Samples of the road dataset: (a) a dry road, (b) a wet road, (c) a snowy road, (d) a muddy road, and (e) other roads.

Table 1
Architecture and parameters of the proposed CNN model.

Layer	Type	Feature map	Step	Convolution kernel	Pooling layer
1	Input	$127 \times 127 \times 3$	–	–	–
2	Convolution	$32@127 \times 64$	1×2	$5 \times 5 \times 3$	–
3	Normalization	$32@127 \times 64$	–	–	–
4	Pooling	$32@127 \times 64$	1×1	–	2×2
5	Convolution	$64@127 \times 32$	1×2	$3 \times 3 \times 32$	–
6	Normalization	$64@127 \times 32$	–	–	–
7	Pooling	$64@127 \times 32$	1×1	–	2×2
8	Convolution	$128@127 \times 16$	1×2	$1 \times 1 \times 64$	–
9	Normalization	$128@127 \times 16$	–	–	–
10	Pooling	$128@127 \times 16$	1×1	–	2×2
11	Fully-connected	$128@1 \times 1$	–	–	–
12	Fully-connected	$2048@1 \times 1$	–	–	–
13	Output	$5@1 \times 1$	–	–	–

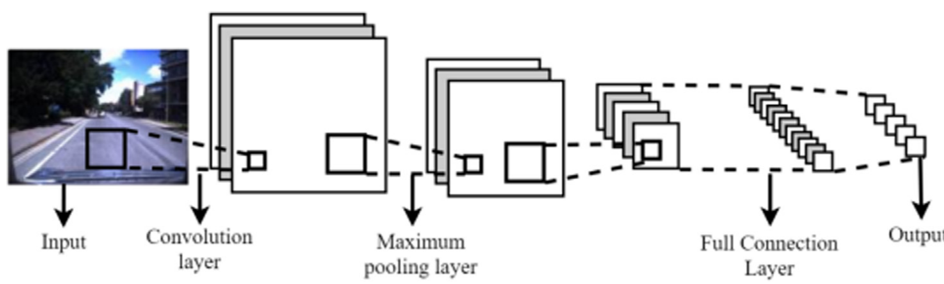


Fig. 2. A schematic diagram of the proposed CNN model for road surface classification.

$$f(x) = \begin{cases} x, & x \geqslant 0 \\ 0, & x < 0 \end{cases} \tag{1}$$

This function is represented by one threshold value and hence fast network convergence can be achieved using this function. However, ReLU neurons can easily be lost in the training process because of frequent changes to the weights. Specifically, if a large gradient is applied to the input of a ReLU unit and the weight changes so that the unit input is less than 0 and its output is 0, the ReLU unit will not resume the operation for any subsequent data points.

In order to avoid data leakage for the negative half-axis part of the ReLU activation function, a leakage value is added to the negative half-axis interval of the ReLU function. The modified function is called the Leaky-ReLU function, which is mathematically defined as:

$$f(x) = \begin{cases} x, & x \geqslant 0 \\ x/a, & x < 0 \end{cases} \tag{2}$$

In Eq. (2), x represents the input of the activation function, and a is the leakage hyper-parameter that can be learned by the back propagation algorithm.

The SoftSign activation function [21] is softly-saturated in the negative half-axis. While this function is similar to the hyperbolic tangent function, it has a flatter curve, a slower descent derivative, and an output value between −1 and 1. The mathematical form of this function is:

$$f(x) = x/(1 + abs(x)) \tag{3}$$

Alternatively, based on the advantages of the leaky ReLU and SoftSign activation functions, this paper proposes an improved ReLU activation function, which is mathematically defined as follows:

$$f(x) = \begin{cases} \ln(x+1) + ax, & x \geqslant 0 \\ ax/(1 + abs(ax)), & x < 0 \end{cases} \tag{4}$$

In this equation, a is a variable superparameter. When $a = 0$, the form of this function is similar to that of the ReLU function. When $a < 0$, both $f(x \geqslant 0)$ and $f(x < 0)$ have zero values. We denote the proposed activation function by *Gai-ReLU*. The shape of this function is shown in Fig. 3.

Fig. 3 shows that, for the positive half-axis, the ReLU function curve coincides with the leaky ReLU function curve, and the Gai-ReLU function curve lies between those of the ReLU and SoftSign functions. This ensures that the Gai-ReLU activation function can model nonlinearities and also lead to faster convergence. In the negative half-axis region, the ReLU function value is constant and equal to 0, while the leaky ReLU function has a larger slope. This means that both functions have adverse effects on data rectification, while the curve of the SoftSign function is flatter, and the Gai-ReLU function is more easily saturated.

The partial derivatives of the Gai-ReLU function are as follows:

$$\frac{\partial f(x)}{\partial x} = \begin{cases} 1/(x+1) + a, & x \geqslant 0 \\ a/(1 + abs(ax))^2, & x < 0 \end{cases} \tag{5}$$

When a neuron is in the excitation region, its gradient decreases with the increase of x and converges to a. When the neuron is in the inhibition region, the gradient will be a small, non-zero value, similar to the leaky ReLU activation function [22]. The global gradient of the Gai-ReLU activation function is more similar to the natural gradient. This ensures speed and accuracy of model learning.

4. Experimental results and analysis

4.1. Experimental setup and datasets

The experiments were performed using a Windows 10 (x64) operating system. The software platform is MATLAB R2018b, running on a GeForce GTX 880M GPU workstation. A single network

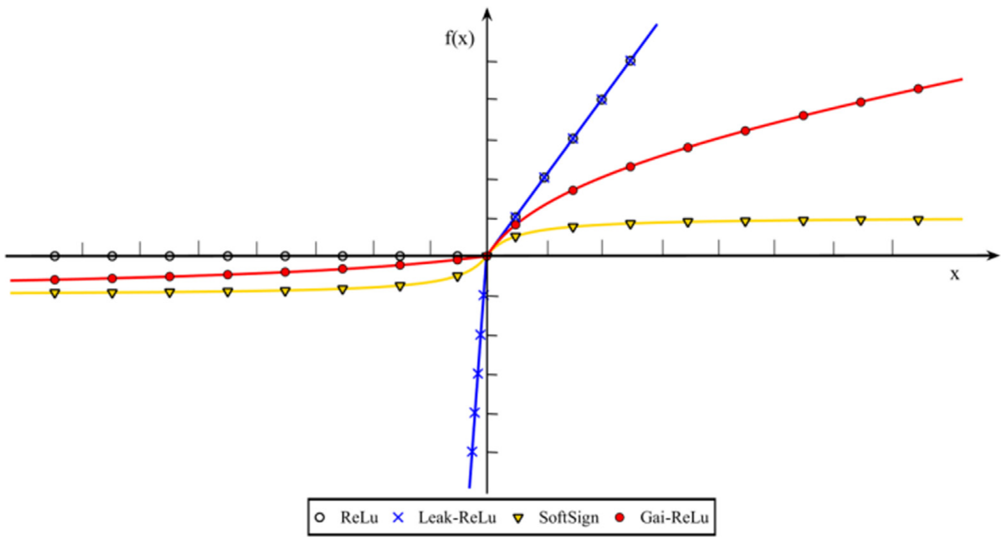


Fig. 3. The mathematical shapes of four types of activation functions.

completes training in approximately 4.2 h, and classifies a test image in around 0.2 s.

In this paper, the Oxford RobotCar Dataset and the KITTI Dataset are used to train and test deep learning models for classifying road surface conditions into five categories: a dry road, a wet road, a snow-covered road, a muddy road and other roads. After image preprocessing, each category has 1600 training images and 400 test images.

A flowchart of the experiments is shown in Fig. 4. The front of the data collection vehicle is selected as the symmetrical center base plane, and the front 127×127 rectangular pixels are the region of interest. After gathering the road surface condition data, the following image preprocessing steps were applied: removing blurred images, applying homomorphic filtering to unify the illumination intensities of the images [23], attaching image labels to the images, and feeding the data to the deep learning model for training and testing.

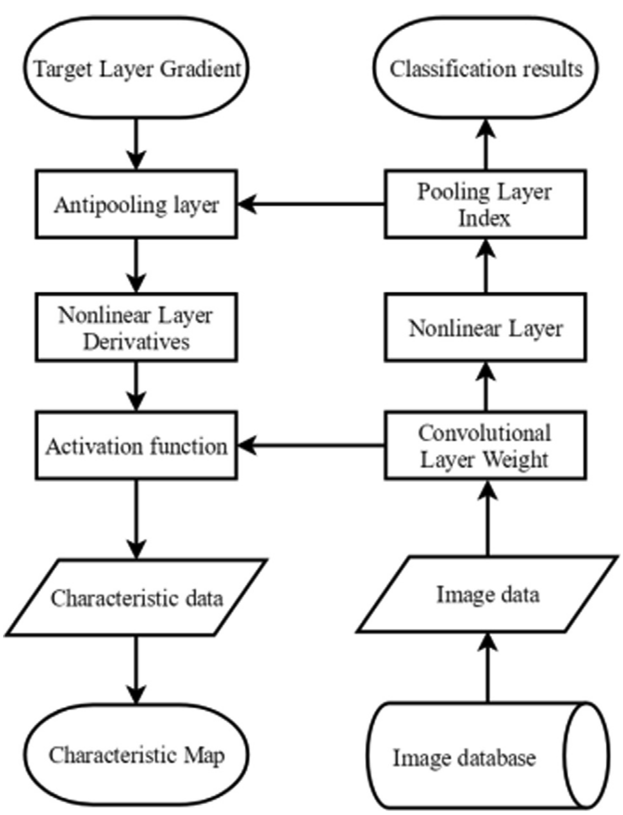


Fig. 4. A flowchart of the proposed road condition classification system.

For all experiments, the model accuracy is measured by the test accuracy of the collected dataset, which is given by

$$accuracy = \frac{1}{m}\sum_{i=1}^{m} I\left\{\hat{y}_i = y_i\right\} \tag{6}$$

where $\hat{y}_i$ is the predicted category of the i sample, y_i is the real category of the i sample, and m is the number of samples.

4.2. Performance comparison for different activation functions

In this paper, deep learning models equipped with the Gai-ReLU function are compared against models with the functions of TanH, SoftMax, ReLU and leaky ReLU, respectively. In addition, comparison is made as well against the results of two traditional machine learning models, namely support vector machines (SVM) and back-propagation (BP) neural networks. The results are shown in Fig. 5. For each activation function, the value of the hyperparameter α is initialized to 0.01. The SVM and BP models adopt HSV color-space features and gray-level co-occurrence matrix (GLCM) features. The classification accuracy for each model is shown in Table 2.

Table 2 and Fig. 5 show the classification results for the aforementioned eight machine learning models on the road datasets. The deep learning models with the Gai-ReLU, ReLU, and leaky ReLU activation functions achieve accuracies exceeding 90%, and hence demonstrate the superiority of the deep learning approaches. The classifier with the TanH function produces the worst accuracy due to the limitations of this function. Traditional machine learning methods like the SVM and the BP neural network produce inferior results, as the hand-crafted features of these classifiers are not suitable to represent light intensity changes. The SVM classifier is suitable for small datasets (about 500 images), and this is why it does not achieve good performance on the road datasets. The deep learning models with the ReLU and leaky ReLU functions have almost the same accuracy. In fact, the classifier with the leaky ReLU function is slightly better than that with the ReLU function. This

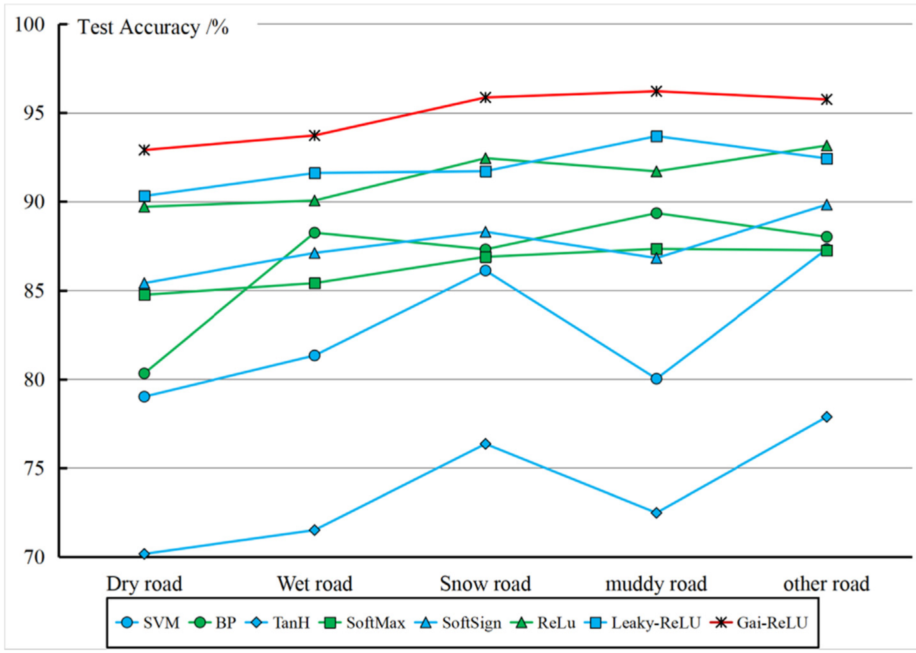


Fig. 5. A comparison of the experimental accuracies of different learning models for the five categories of road surfaces.

Table 2
Classification accuracy for each learning model.

Activation function	SVM	BP	TanH	SoftMax	SoftSign	ReLU	Leaky-ReLU	Gai-ReLU
Training accuracy %	84.35	85.18	78.10	86.56	88.22	92.17	92.46	95.48
Test accuracy %	82.77	86.65	73.68	86.30	87.49	91.41	91.95	94.89

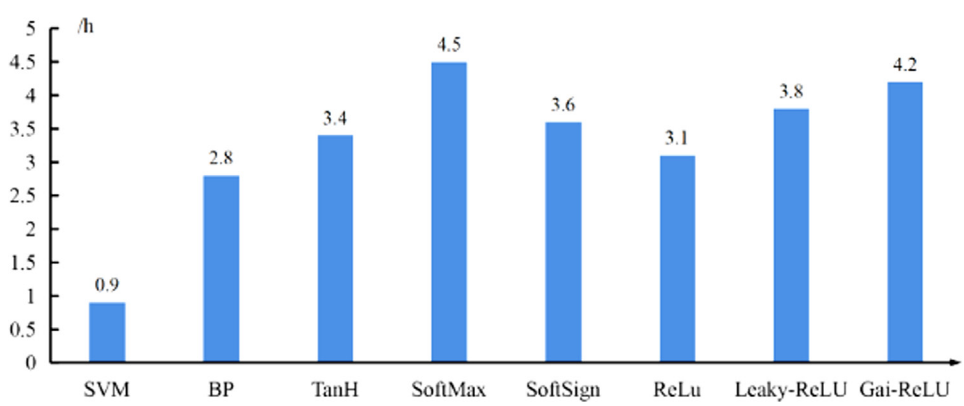


Fig. 6. Training times for different learning models.

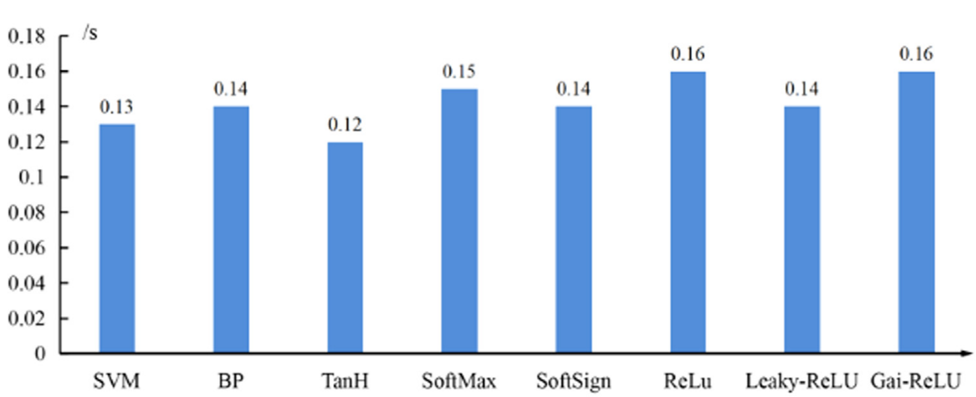


Fig. 7. Average time spent classifying an image with different learning models.

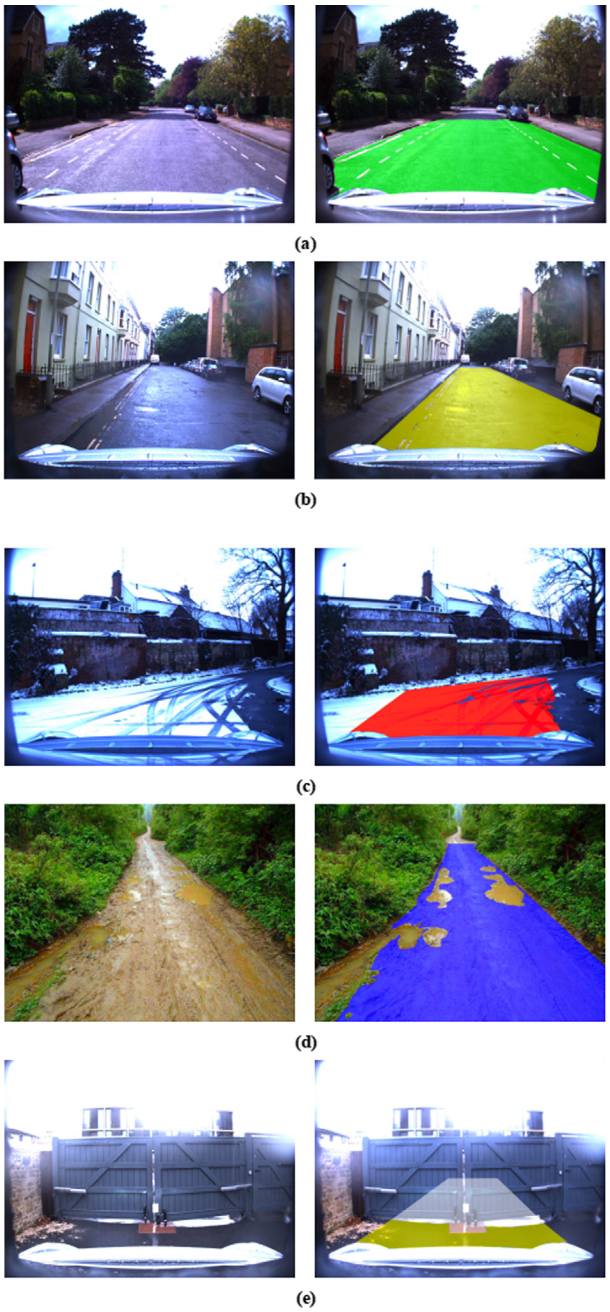


Fig. 8. A comparison of the contrast maps for the five categories of road surface conditions: (a) a dry road, (b) a wet road, (c) a snowy road, (d) a muddy road, and (e) other roads.

shows that the negative half-axis of the activation function has little influence on the accuracy. The test accuracy of the classifier with the Gai-ReLU function is 94.89%, which is 3.48% higher than that with the ReLU function. This confirms the relative importance of the Gai-ReLU activation function proposed in this paper.

By using the available datasets to conduct model training for the models above, the training time for each model can be obtained. As shown in Fig. 6, the computational complexity of the activation functions has an important impact on model training times. Indeed, the training time of the SoftMax activation function model is the largest, while the training time for the ReLU function model is the smallest because of the relative simplicity of this function. For training the SVM model, the RGB, HSV, and surface roughness features are extracted from the training images. The training times of the SVM model are the smallest due to the simplicity of those features. Fig. 7 shows the average time taken by each learning model to classify a 1280×960 test image. It can be seen that the test times of all models for a given road image are almost the same.

In practical applications, the accuracy and time consumption of the classifiers of images of road conditions are typically studied and analyzed. However, our deep learning model is pre-trained, and so its training time cannot be fully quantified. The training of the Gai-ReLU activation function model requires 4.2 h, and it takes 0.16 s to classify a test image.

Fig. 8 shows the results of classifying images of different road conditions using the Gai-ReLU activation function model in MATLAB. After setting the area to be identified in the test image, different colors are used to cover different road conditions: dry roads in green, wet roads in yellow, snowy roads in red, muddy roads in blue, and other roads in white. From Fig. 8, we can see that the classification performance for road conditions is good, and can be applied in the classification of actual road condition images.

4.3. Comparative analysis on general datasets

To show the wide applicability of the Gai-ReLU activation function model, its performance is compared against other learning models on general datasets other than the road datasets. The CIFAR-10 dataset [24] was selected as the general dataset for this setup. Five categories of images were considered: aircraft, automobile, bird, horse and boat. Each image has 32×32 pixels, where each category comprises 1600 training images and 400 testing images. The collected images are shown in Table 3.

From Table 4, it can be seen that the accuracy of the Gai-ReLU based model is clearly higher than that of the TanH and SoftMax functions on the general datasets. Compared with the ReLU model, the Gai-ReLU model accuracy is improved by 3.02%. In this experiment, the SVM and BP methods have very poor accuracy, since the features were not tailored to the data. The advantages of explicit

Table 3
CIFAR-10 database images.

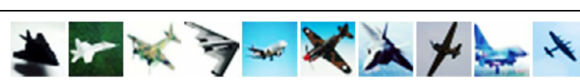
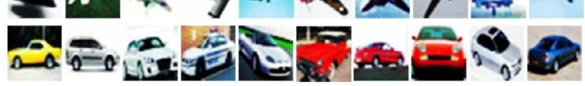
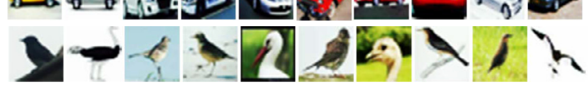
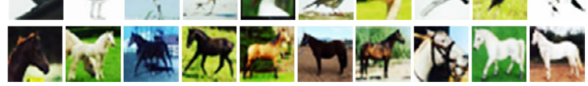
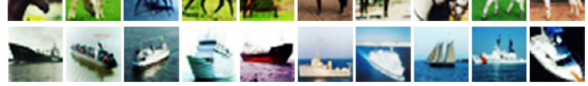
Aircraft	
Automobile	
Bird	
Horse	
Boat	

Table 4
Classification results for different activation functions using general datasets.

Activation function	SVM	BP	TanH	SoftMax	SoftSign	ReLU	Leaky ReLU	Gai-ReLU
Training accuracy/%	55.74	59.33	77.47	86.41	87.55	92.39	94.40	96.21
Test accuracy/%	53.42	56.90	75.66	84.45	86.79	90.91	91.26	93.93

in-depth learning is being unsupervised and that it does not need to extract features manually. However, these advantages are not obvious when comparing the ReLU and leaky ReLU activation function models, which appear to have no clear relative advantages to each other. In conclusion, the superiority of the deep learning model with the proposed Gai-ReLU activation function has been demonstrated, together with its excellent generalization performance.

5. Conclusions

Deep convolutional neural networks have powerful recognition and learning abilities. These networks produce excellent performance in solving image classification and recognition problems. In this paper, a deep learning method, combined with an improved activation function and a GPU with high computing power, is used to classify five common types of road surface conditions. The results show that the deep model equipped with a Gai-ReLU activation function has a high classification accuracy of 94.89%. Moreover, the deep learning model with this activation function has good generalization and applicability to real-world applications. Our experiments show that the proposed deep learning model based on an improved activation function still has room for improvement with respect to hyperparameters and structure. In the next stage, the focus of our work will be to improve the recognition accuracy, consummate the road surface condition dataset, and improve the network structure and hyperparameters.

Declaration of Competing Interest

The authors declare that they have no conflicts of interest.

Acknowledgements

The work described in this paper was supported by the National Natural Science Foundation of China (Grant No. 51205246, 51775328).

References

[1] T.N. Wiesel, D.H. Hubel, Effects of visual deprivation on morphology and physiology of cells in the cat's lateral geniculate body, J. Neurophysiol. 26 (6) (1963) 978–993.
[2] J.M. Keller, D. Liu, D.B. Fogel, Multilayer neural networks and backpropagation, in: Fundamentals of Computational Intelligence: Neural Networks, Fuzzy Systems, and Evolutionary Computation, Spring Science+Business media B.V, New York, 2017, p. 378.
[3] Y.L. Lecun, L. Bottou, Y. Bengio, et al., Gradient-based learning applied to document recognition, Proc. IEEE 86 (11) (1998) 2278–2324.
[4] Y.C. Wu, F. Yin, C.L. Liu, Improving handwritten Chinese text recognition using neural network language models and convolutional neural network shape models, Pattern Recogn. 65 (2017) 251–264.
[5] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, Commun. ACM 60 (06) (2017) 84–90.
[6] Y. Fu, C. Aldrich, Flotation froth image recognition with convolutional neural networks, Miner. Eng. 132 (8) (2019) 183–190.
[7] K. Shi, H. Bao, N. Ma, Forward vehicle detection based on incremental learning and fast R-CNN, 2017 13th International Conference on Computational Intelligence and Security (CIS), IEEE, 2017.
[8] D.U. Lan, L.I.U. Bin, W.A.N.G. Yan, et al., Target detection method based on convolutional neural network for SAR image, J. Electr. Inform. Technol. 38 (12) (2016) 3018–3025.
[9] Z. Huina, L.I. Yumei, F.U. Yingying, et al., Research on natural scene image classification based on Haar-CNN Model, J. Sichuan Norm. Univ. 40 (01) (2017) 119–126.
[10] A. Kuehnle, W. Burghout, Image-based winter road condition recognition, Applications of Advanced Technologies in Transportation, ASCE, 1998.
[11] Z. Liu, J. Yang, C. Liu, Extracting multiple features in the CID color space for face recognition, IEEE Trans. Image Process. 19 (9) (2010) 2502–2509.
[12] S. Zhang, S. Zhao, Y. Sui, et al., Single object tracking with fuzzy least squares support vector machine, IEEE Trans. Image Process. 24 (12) (2015) 5723–5738.
[13] Liu Yuexia, Cheng Yunfei, Wu Wang, A survey of the application of deep learning in computer vision, Global Intelligent Industry Conference, GIIC, 2018.
[14] D. Erhan, Y. Bengio, A.C. Courville, et al., Why does unsupervised pre-training help deep learning?, J Mach. Learn. Res. 11 (3) (2010) 625–660.
[15] P. Philipp, V. Felix, Optimal approximation of piecewise smooth functions using deep ReLU neural networks, Neural Networks (2017), arXiv:1709.05289.
[16] Fréderic Godin, J. Degrave, J. Dambre, et al., Dual rectified linear units (DReLUs): a replacement for tanh activation functions in quasi-recurrent neural networks, 2017. arXiv:1707.08214.
[17] D.R. Harvey, RELU special issue: editorial reflections, J. Agric. Econ. 57 (2) (2006) 8.
[18] W. Maddern, G. Pascoe, C. Linegar, et al., 1 year, 1000 km: the oxford robotcar dataset, Inter. J. Robot. Res. 36 (01) (2017) 3–15.
[19] A. Geiger, P. Lenz, C. Stiller, et al., Vision meets robotics: The KITTI dataset, Inter. J. Robot. Res. 32 (11) (2013) 1231–1237.
[20] G. Hinton, S. Osindero, M. Welling, et al., Unsupervised discovery of nonlinear structure using contrastive backpropagation, Cognitive Science 30 (4) (2006) 725–731.
[21] H. Zhao, F. Liu, L. Li, et al., A novel softplus linear unit for deep convolutional neural networks, Applied Intelligence 48 (7) (2018) 1707–1720.
[22] Hari Krishna Vydana, Anil Kumar Vuppala, Investigative study of various activation functions for speech recognition, Communications (NCC), 2017 Twenty-third National Conference on, 2017.
[23] Y. Yu, X.B. Li, H.Y. Sun, Color images illumination compensation method for human face region detection, Appl. Mech. Mater. 678 (2014) 162–165.
[24] A. Krizhevsky, G. Hinton, Learning multiple layers of features from tiny images [DB], Computer Science Department, University of Toronto, 2009.